Proof, Sets, and Logic

M. Randall Holmes

November 30, 2012

For Jonathan

Contents

1 Introduction

1.1 Version Notes

These are notes to myself as the editor of the document. I will highlight changes which actually affect material currently being lectured (or past material), which will of course also be of interest to current students.

November 30, 2012: reading this and preparing updates as part of my sabbatical work. Noted the requests in the text for elementary results about Boolean algebra and applications in ordinary mathematics of Zorn's lemma.

I have written a section defining NFU and proving its consistency and am now working on a section investigating the mathematics of NFU (meaning its mathematical power over and above what it inherits from TSTU). This section mentions the fact that we can bootstrap from TSTU foundations to NFU foundations (not via Zermelo-style set theory).

June 7, 2011: Finished reading through the document preparatory to extending it. Minor typos corrected.

Material about ultrafilters and ultrapowers added.

June 6, 2011: I am going to work on filling this out this summer.

This text needs insertions about Marcel if it is to be used with it.

First decision: insert the use of the absurd $\perp$.

p. 40 note on elementary theorems about set operations should be cashed out. p. 116 note on general applications of Zorn's Lemma should be cashed out

July 19, 2008: Considerably modified the outline of the book. The alternative set theories to be covered are restricted to the stratified theories with universal set. The basic results on ambiguity are present in draft (the cases of theories with unbounded and bounded types are treated differently). [this was actually done before July 19 but I didn't update the version notes or post to the web].

Added material on forcing in type theory. Note that I have basic constructions for consistency and independence set up in the outline in to

be in the first instance in type theory rather than in either untyped or stratified theories. My foundation for mathematics as far as possible is in TSTU here, not in either ZFC or NFU.

July 14, 2008: Massively rewrote the outline of the unrealized last section on stratified set theories.

I saved the previous version as proofsetslogicjuly14.tex then did major surgery on section 4.

The section now contains an initial account of Zermelo's axioms in their original form, followed by a justification of the axioms in terms of an interpretation in the theory of set pictures in type theory (which immediately precedes it in the text), together with the observation that the Axiom of Rank holds in this interpretation. After that comes a new section on implementation of mathematical concepts in Zermelo set theory with the Axiom of Rank. The Scott trick is introduced; the von Neumann definitions are also given.

Material in the old version about an axiomatic cumulative type theory is removed. There may still be unintelligible references to it which need to be removed. Some older material which still seems useful is retained, but there is now a fair amount of redundancy and rearrangements of the material will surely be needed.

Some comments: I may want to do some forcing and some Frankel-Mostowski methods in type theory. It is useful to be able to do basic consistency and independence results in the basic foundational theory, and it may make it easier to follow the development in NFU or NF. Definitely introduce TNTU (and TNT specifically) as examples in the logic section. Methods in type theory which are portable to NF(U) are neatly summarized as those which work in TNT (don't do recursive constructions on types; keep things ambiguous). Something about my reservations about ZFC would not be out of place here somewhere.

July 13, 2008: Filled in lots of proofs in the section on well-founded extensional relations. Cleaned up various errors and infelicities. My proof of extensional collapse is not ideal (I should say something better about uniqueness).

Correcting an omission in the definition of "membership" E for weak membership diagrams. I didn't provide an implementation of atoms

in weak set pictures: double singletons of the atoms of the original relations work smoothly (and preserve the basic theorems) but one needs to modify the definitions of weak set picture and of the relation E on weak set pictures.

July 12, 2009: I cleaned up the section on isomorphism types of well-founded extensional relations (it is now a single section, not Old Version and New Version). Some of the stated Theorems will have proofs added (some of the proofs were present in versions now erased). The new development has the nice (and unexpected) feature that it supports development of ZFA (or more accurately Mac Lane with atoms) essentially as smoothly as it supports extensional Mac Lane, which is very convenient given our philosophical commitment to weak extensionality already stated. This section needs to be fleshed out (with proofs and perhaps additional discussion and examples) then the following Section 4 on untyped set theory needs to be passed through with the idea that its relation to the theory of well-founded extensional relations should be exploited at every turn.

Notice that the admission of urelements gets even stronger positive press because it prevents the truncation of the construction of the cumulative hierarchy of isomorphism types of well-founded extensional relations (the world of the usual set theory).

Further, it is interesting to note that the von Neumann definition of ordinal numbers is *entirely* natural if the world of untyped set theory is viewed as being constructed from the isomorphism types of well-founded extensional relations. The von Neumann naturals (as untyped sets) are the isomorphism classes of strict well-orderings. This gets scrambled a bit when one passes to NFU, but this is also instructive: the ordinals of untyped set theory are restricted to the standard world in a way that the ordinals of NFU are not.

I think I might want to put a section on independence of the Axiom of Choice from type theory in the first part. That would also provide a natural platform for cautions about the difference between choiceful and choice-free mathematics, which are needed if we are finally to discuss the weirdness of NF intelligibly.

July 10, 2009: Continuing to work. Added characterization of the order type of the rationals. Added Hartogs and Sierpinski theorems. Those

might need some cleanup. Tried to tighten up the discussion which motivates TSTU by consideration of one's desire to be able to extend the $\sqsupset$ operation. It is certainly a valid point which is being made, but I would like to say it neatly. It is nice to get to it before one tries to construct NF(U); thus one does not naturally wander into the NF problem.

July 9, 2009: I am editing the text again, much later, with an eye to starting the addition of the last section(s) in which NFU and related issues will make their appearance. Since I am no longer addressing current students as readers, I erased the old version notes.

Recent changes: I eliminated many but not all notes introduced by "NOTE:" in the text (notably ones that had actually already been dealt with). I cleaned up the format of definitions through a large part of the document. I made superscripts indicating iteration of type-shifting operations boldface, like type indices, though I am not sure that this was appropriate in all parts of the text.

I wonder if there is a place (perhaps in the development of Zermelo set theory) to note the complication of the ADT of the ordered pair revealed by Adrian Mathias. In the type theory section there is no reason to bring this up: any list constructor IS an ordered-pair-for-building-relations in type theory. Maybe that is an observation worth making too.

The discussion of well-founded extensional relations needs to be polished up and unified and at least the option of using it to introduce untyped set theory needs to be clearly indicated.

I might want to give a more thorough discussion of the intrinsic mathematical reasons to prefer *TSTU* over *TST* which are revealed in the theory of cardinal number before *NFU* is ever considered. These are laid out: the point is that quite reasonable views about the ability to iterate the exp operation on cardinals are not supported in *TST*. *TSTU* neatly solves the problem. In fact, there might be a more general thing worth saying: any construction we can define in type theory building on top of type 0 we probably ought to be able to copy down into urelements of type 1. So the rate at which the types increase in size should exceed anything we can define. Ideas for reflection principles naturally follow.

I should look at the logical material for my recent M387 and see if some of it has a place here.

1.2 Introductory Remarks

This is being written as a textbook for Math 502, Logic and Set Theory, at Boise State University, on the practical level.

On the Platonic level, this is intended to communicate something about proof, sets, and logic. It is about the foundations of mathematics, a subject which results when mathematicians examine the subject matter and the practice of their own subject very carefully.

The "proof" part refers to an informal discussion of mathematical practice (not all *that* informal) which will serve as a springboard for the "logic" component. It also introduces formal notation for logic.

The "sets" part refers to a careful development of mathematical ontology (a fancy word for "what are we talking about"?): familiar mathematical concepts are analyzed in terms of the fundamental concept of a *set*. This section gives us an opportunity to practice the proof skills of which section 1 provides an overview. A distinctive feature of our development is that we first develop basic concepts of set theory in a typed theory of sets, then make the transition to the more usual untyped set theory in a separate section.

The "logic" part refers to a much more formal discussion of how we prove things, which requires both the "proof" and "sets" components to work properly, and in which bits of language (sentences and noun phrases) and proofs are actually mathematical objects.

All of this is supported by some software: the formal logic introduced in section 5 (and one of the alternative set theories introduced in section 6) are the logic of our sequent theorem prover Marcel, to which we will have occasion to refer, and which will be used for some lab exercises. We hope to find that experience with Marcel will assist the learning of formal logic.

The final section on alternative set theories will probably not be reached in the course (or in a first course, at any rate) but has some bearing on other ways we could get from type theory to set theory and on the way set theory is implemented in Marcel.

2 Proof

In this section we discuss how we make "formal proofs" (really, as we will see in the Logic section, rather *informal* proofs) in English, augmented with formal notation.

Our framework is this. We will identify basic logical structures of statements. Statements have two fundamental roles in proofs which need to be carefully distinguished: there are statements which we are trying to deduce from our current assumptions, which we will call "goals", and there are statements already assumed or deduced from the current assumptions which we are allowed to use, which we will call "posits". The reason we call these last "posits" instead of something like "theorems" or "conclusions" is that posits may be consequences of statements which we have only assumed for the sake of argument: a posit is not necessarily a theorem. For each basic logical structure, we will indicate strategies for deducing a goal of that form (from the currently given posits) and strategies for using a posit of that logical form to deduce further consequences. Further, we will supply formal notation for each of the basic logical structures, and we will say something about the quite different English forms which statements of the same underlying logical form may take.

It is useful to note that my use of the word "posit" is eccentric; this is not standard terminology. We can adopt as a posit any current assumption, any previously proved theorem, or anything which follows logically from current assumptions and theorems. We allow use of "posit" as a verb: when we adopt A as a posit, we posit A (to posit is either to *assume* for the sake of argument or to *deduce* from previous posits).

We are trying to say carefully "deduce" rather than "prove" most of the time: what we can *prove* is what we can deduce without making any assumptions for the sake of argument.

2.1 Basic Sentences

Mathematical sentences (being sentences of natural language) have subjects, verbs and objects. Sentences in formal mathematical language have similar characteristics. A typical mathematical sentence already familiar to you is $x < y$ (though we will see below that we will usually call this particular (grammatical) sentence a "formula" and not a "sentence" when we are being technical). Here x and y are noun phrases (the use of letters in mathematical

notation is most analogous to the use of *pronouns* in English, except that for precision of reference mathematical language has a lot more of them). $<$ is the verb, in this case a transitive verb with subject and object. In the parlance of mathematical logic, a transitive verb is called a "binary predicate". Another typical kind of mathematical sentence is "x is prime". Here the verb phrase "is prime" is viewed as an intransitive verb (we don't distinguish between adjectives and intransitive verbs as English does). We can't think of examples of the use of intransitive verbs in mathematical English, though we are sure that they do exist. An adjective or intransitive verb is a "unary predicate" in mathematical logic. Two commonly used words in mathematical logic which have grammatical meanings are "term" and "formula": a "term" is a noun phrase (for the moment, the only terms we have are variables, but more term constructions will be introduced as we go on) and a "formula" is a sentence in the grammatical sense ("sentence" in mathematical logic is usually reserved for formulas not containing essential references to variables: so for example $x < y$ is a formula and not (in the technical sense) a sentence, because its meaning depends on the reference chosen for x and y, while $2 < 3$ is a formula and a sentence (no variables) and $(\exists x.x < 2)$ is a formula and a sentence (the x is a dummy variable here)). What we call "basic sentences" (using terminology from grammar) in the title of this section will really be called "atomic formulas" hereinafter.

The English word "is" is tricky. In addition to its purely formal use in "x is prime", converting an adjective to a verb phrase, it is also used as a genuine transitive verb in formulas like "x is the square of y", written $x = y^2$ in mathematical language. The $=$ of equality is a transitive verb (as far as we are concerned: it is not treated the same by English grammar) and also part of our basic logical machinery.

The English word "is" may signal the presence of another binary predicate. A formula like "x is a real number" may translate to $x \in \mathbb{R}$, where $\in$ is the predicate of *membership* and $\mathbb{R}$ is the name of the set of all real numbers. For that matter, the formula "x is prime" could be read $x \in \mathbb{P}$ where $\mathbb{P}$ is here supposed to be the set of all prime numbers.

In our formal language, we use lower case letters as variables (pronouns). There will be much more on the care and feeding of variables later on. Some special names for specific objects will be introduced as we go on (and in some contexts lower case letters (usually from the beginning of the alphabet) may be understood as names (constants)). Capital letters will be used for predicates. $P(x)$ ("x is P") is the form of the unary predicate formula.

$x\,R\,y$ is the form of the binary predicate formula. Predicates of higher arity could be considered but are not actually needed: a ternary predicate formula might be written $P(x, y, z)$. The specific binary predicates of equality and membership are provided: $x = y$, $x \in y$ are sample formulas. Much more will be heard of these predicates later.

We will have another use for capital letters, mostly if not entirely in this Proof part: we will also use them as variables standing for sentences. We use variables A, B, C for completely arbitrary sentences (which may in fact have complex internal structure). We use variables P, Q, R for propositions with no internal structure (atomic formulas). Once we get to the sections on set theory we will once again allow the use of capital letters as variables representing objects (usually sets).

2.2 Conjunction

This brief section will review the mathematical uses of the simple English word "and". The use of "and" as a conjunction to link sentences is what is considered here. If S is "snow is white" and G is "grass is green", we all know what "snow is white and grass is green" means, and we formally write $S \wedge G$.

Certain English uses of "and" are excluded. The use of "and" to link noun phrases as in "John and Mary like chocolate" is not supported in mathematical language. This use does have a close connection to the logical "and": the sentence is equivalent to "John likes chocolate and Mary likes chocolate". One should further be warned that there is a further complex of uses of "and": "John and Mary went out together" does not admit the logical analysis just given, nor (probably) does "John and Mary moved the half-ton safe". There is an example of the nonlogical use of "and" in mathematical parlance: there is a strong tempation to say that the union of two sets a and b, $a \cup b$, consists of "the elements of a and the elements of b". But $x \in a \cup b$ is true just in case $x \in a$ *or* $x \in b$. Another example of a use of "and" which is not a use of $\wedge$ is found in "x and y are relatively prime".

We note and will use the common mathematical convention whereby $t\,R\,u\,S\,v$ is read $t\,R\,u \wedge u\,S\,v$, as in common expressions like $x = y = z$ or $2 < 3 \leq 4$. This chaining can be iterated:

$$t_0\,R_1\,t_1\,R_1\,t_2\,\ldots\,t_{n-1}\,R_n\,t_n$$

can be read

$$t_0 \, R_1 \, t_1 \wedge t_1 \, R_2 \, t_2 \wedge \ldots \wedge t_{n-1} \, R_n \, t_n.$$

Proof Strategy: To deduce a goal of the form $A \wedge B$, first deduce the goal A, then deduce the goal B.

If you have posited (assumed or deduced from current assumptions) $A \wedge B$, then you may deduce A and you may deduce B.

The operation on propositions represented by $\wedge$ is called *conjunction*: this is related to but should not be confused with the grammatical use of "conjunction" for all members of the part of speech to which "and" belongs.

2.3 Disjunction

This subsection is about the English word "or".

Again, we only consider "or" in its role as a conjunction linking sentences; the use of "or" in English to construct noun phrases has no analogue in our formal language.

When we say "A or B" in mathematics, we mean that A is true or B is true *or both*. Here we draw a distinction between senses of the word "or" which is also made formally by lawyers: our mathematical "or" is the "and/or" of legal documents. The (presumably) exclusive or of "You may have chocolate ice cream or you may have vanilla ice cream" is also a logical operation of some interest but it is not yet introduced here.

We write "A or B" as $A \vee B$, where A and B are sentences.

Proof Strategy: To deduce a goal $A \vee B$, deduce A. To deduce a goal $A \vee B$, deduce B. These are two different strategies. We will see below that two more powerful strategies exist (generalizing these two): To deduce a goal $A \vee B$, assume $\neg A$ ("not A") and deduce B; To deduce a goal $A \vee B$, assume $\neg B$ and deduce A.

For a fuller discussion of this kind of proof strategy which involves the introduction of an additional assumption, see the subsection on implication below (and for more about negation see the section on negation below).

To use a posit $A \vee B$ (assumed or deduced from the current assumptions) to deduce a conclusion G, we use the strategy of *proof by cases*: first

deduce G from the current assumptions with A replacing $A \vee B$, then deduce G from the current assumptions with B replacing $A \vee B$ [both of these proofs are needed].

The operation on propositions represented by $\vee$ is called *disjunction*.

2.4 Implication

The sentences "if A, then B", "B if A", "(that) A (is true) implies (that) B (is true)" all stand for the same logical construction. Other, specifically mathematical forms of the same construction are "(that) A (is true) is sufficient for B (to be true)" and "(that) B (is true) is necessary for A (to be true)". We provide optional padding phrases in parentheses which are needed in formal English because a proposition cannot grammatically live in the place of a noun phrase in an English sentence. Our formal notation for any of these is $A \to B$.

Don't spend a lot of time worrying about "necessary" vs. "sufficient" for purposes of reading this text – I only occasionally use them. But other writers use them more often; if you are going to read a lot of mathematics you need to know this vocabulary.

It is important to notice that unlike previous constructions this one is not symmetrical: "if A, then B" is not equivalent to "if B, then A".

Proof Strategy: To deduce a goal $A \to B$, assume A (along with any other assumptions or previously reduced results already given in the context) and deduce the goal B. Once the goal B is proved, one withdraws the assumption that A (it is local to this part of the proof). The same remarks apply to the negative assumptions introduced in the more general strategy for proving disjunctions indicated above. An alternative strategy for proving $A \to B$ (called "proving the contrapositive") is justified in the section on negation: assume $\neg B$ and adopt $\neg A$ as the new goal.

A posit of the form $A \to B$ is used together with other posits: if we have posited $A \to B$ and we have also posited A, we can deduce B (this rule has the classical name *modus ponens*). We will see below that we can use posits $A \to B$ and $\neg B$ to deduce $\neg A$ as well (the rule of *modus tollens*).

Another way to think of this: if we have a posit $A \to B$ we can then introduce a new goal A, and once this goal is proved we can deduce the further conclusion B. [or, following the pattern of *modus tollens*, we can introduce a new goal $\neg B$, and once this goal is proved we can deduce $\neg A$].

The operation on propositions represented by $\to$ is called *implication*.

The additional strategies indicated in this section and the section on disjunction which involve negation ($\neg$) will be further discussed in the section on negation below.

2.5 Biconditional and Exclusive Or

When we say "A if and only if B", "A (being true) is equivalent to B (being true)", "A exactly if B", or similar things we are saying that A and B are basically the same statement. Formal notations for this is $A \leftrightarrow B$. We have often used $\equiv$ for this operator elsewhere, and the notation of Marcel (==) is motivated by this alternative notation. "A iff B" is a learned abbreviation for "A if and only if B" which is used in mathematical English.

Proof Strategy: To deduce a goal of the form $A \leftrightarrow B$, deduce $A \to B$ and deduce $B \to A$. Since there are at least two strategies for deducing these implications, there are a number of ways to structure the proof.

One can use a posit of the form $A \leftrightarrow B$ in a number of ways. From posits $A \leftrightarrow B$ and A, we can deduce B; from posits $A \leftrightarrow B$ and B we can deduce A. More powerfully, if we have posits $A \leftrightarrow B$ and some complex $C[A]$, we can deduce $C[B]$ (simply replace occurrences of A with B) or symmetrically from posits $A \leftrightarrow B$ and $C[B]$ we can deduce $C[A]$.

The operation represented by $\leftrightarrow$ is called the *biconditional*.

We note without pursuing the details at this point that $A \not\leftrightarrow B$ (another commonly used notation is $A \oplus B$) is our notation for the "exclusive or": A or B is true but not both.

A common format for a theorem is to give a list of statements and assert that all of them are equivalent. A strategy for proving that statements $A_1, \ldots, A_n$ are equivalent is to show that $A_i \to A_{i+1 \bmod n}$ for each appropriate

i (showing that each statement implies the next in a cycle). In a theorem of this type several linked cycles may be present.

We note that $(A \leftrightarrow B) \leftrightarrow C$ is equivalent to $A \leftrightarrow (B \leftrightarrow C)$ but *not* equivalent to $(A \leftrightarrow B) \wedge (B \leftrightarrow C)$.

2.6 Negation and Indirect Proof

It is common to say that the logical operation of negation (the formal notation is $\neg A$) means "not A". But "not A" is not necessarily an English sentence if A is an English sentence. A locution that works is "It is not the case that A", but we do not in fact usually say this in either everyday or mathematical English.

"Not snow is white" is ungrammatical; "It is not the case that snow is white" is pedantic; "Snow isn't white" is what we say. If R is a relation symbol, we will often introduce a new relation symbol $\not R$ and let $x \not R y$ be synonymous with $\neg x R y$. The use of $\neq$ and $\notin$ should already be familiar to the reader.

We do not as a rule negate complex sentences in English. It is possible to say "It is not the case that both A and B are true" but this is only a formal possibility: what we would really say is "A is false or B is false". It is possible to say "It is not the case that either A or B is true" but this is also only a formal possibility: what we would really say is "A is false and B is false". The logical facts underlying these locutions are the identities

$$\neg(A \wedge B) \leftrightarrow (\neg A \vee \neg B)$$

and

$$\neg(A \vee B) \leftrightarrow (\neg A \wedge \neg B),$$

which are known as *de Morgan's laws*. It is pure common sense that we do not need to say "It is not the case that it is not the case that A", when we can so easily say A (the principle of double negation $\neg\neg A \leftrightarrow A$). $\neg(A \rightarrow B) \leftrightarrow A \wedge \neg B$ and $\neg(A \leftrightarrow B) \leftrightarrow (A \not\leftrightarrow B)$ might require a little thought. The former is best approached via the equivalence of $A \rightarrow B$ and $\neg A \vee B$ (which might itself require thought); the result about the negation of $A \rightarrow B$ then follows from de Morgan's laws and double negation. Do please note that we do not here authorize the use of these equivalences as proof strategies (without proof): they are mentioned here only as part of our discussion of the rhetoric of negation in mathematical English!

A statement of the form $A \wedge \neg A$ is called a *contradiction*. It is clear that such statements are always false. It is a logical truth that $A \vee \neg A$ is always true (this is called the law of *excluded middle*).

We introduce the notation $\perp$ for a fixed false statement, which we may call "the absurd".

Proof Strategies:

1: To deduce a goal of the form $\neg A$, add A to your assumptions and deduce $\perp$, the absurd statement. Notice that we will certainly withdraw the assumption A when this proof is done!

2: From A and $\neg A$, deduce $\perp$. The only way to deduce the absurd is from a contradiction.

3: From $\neg\neg A$, deduce A. Otherwise, one can only use a negative hypothesis if the current goal is $\perp$: if we have a posit $\neg A$, use it by adopting A as a goal ("for the sake of a contradiction", so that $\perp$ can be deduced).

The first strategy above is *not* the notorious technique of "proof by contradiction": it is the direct strategy for proving a negative sentence. The strategy of proof by contradiction differs from all our other techniques in being applicable to sentences of any form.

Proof by Contradiction: To deduce any goal A at all, assume $\neg A$ and reason to $\perp$ (by reasoning to a contradiction). Notice that this is the same as a direct proof of the goal $\neg\neg A$.

Principle of Double Negation: $\neg\neg P \leftrightarrow P$

Proof: Part 1 of the proof requires us to deduce P given the assumption $\neg\neg P$: this is given as a basic proof step above. Part 2 requires us to deduce $\neg\neg P$ given the assumption P: to do this, assume $\neg P$ and deduce $\perp$: but this is immediate as we have already assumed P. The proof is complete.

In later parts of the book we will not usually mention $\perp$, so the strategy for proving $\neg A$ will generally be to deduce some contradiction $B \wedge \neg B$ from A (from which the further deduction of $\perp$ is immediate), and the strategy

of proof by contradiction of A will be to deduce some contradiction $B \wedge \neg B$ from $\neg A$ (thus the name).

We prove that $P \to Q$ is equivalent to $\neg Q \to \neg P$. This will give our first extended example of the proof techniques we are advertising.

Contrapositives Theorem: $(P \to Q) \leftrightarrow (\neg Q \to \neg P)$

Proof: This breaks into two subgoals: Goal 1 is to prove $(P \to Q) \to (\neg Q \to \neg P)$ and Goal 2 is to prove $(\neg Q \to \neg P) \to (P \to Q)$.

We prove Goal 1: $(P \to Q) \to (\neg Q \to \neg P)$.

This goal is an implication, so we assume for the sake of argument that $P \to Q$: our new goal is $\neg Q \to \neg P$.

The new goal is also an implication, so we assume $\neg Q$ and have our latest goal as $\neg P$.

To deduce $\neg P$ we need to assume P and deduce $\perp$. We duly assume P. We have already assumed $P \to Q$, so modus ponens allows us to conclude Q. We have already assumed $\neg Q$, so we can conclude $\perp$, which is the goal, which allows us to complete the deduction of our latest goal $\neg P$, and so of the intermediate goal $\neg Q \to \neg P$ and so of Goal 1.

Goal 2 remains to be proved: $(\neg Q \to \neg P) \to (P \to Q)$. To prove this we need to assume $(\neg Q \to \neg P)$ and deduce an intermediate goal $P \to Q$. To deduce this goal, we need to assume P and deduce a second intermediate goal Q. To prove Q, we assume $\neg Q$ and take as our final intermediate goal $\perp$ (this is proof by contradiction). From $\neg Q$ and the earlier assumption $\neg Q \to \neg P$ we can conclude $\neg P$ by modus ponens. From the earlier assumption P and the recently proved $\neg P$ we conclude $\perp$, completing the deductions of all outstanding goals and the proof of the entire theorem.

Notice that we could replace the propositional letters P and Q with any statements A and B, however complex, and the proof above would still work: we have actually proved $(A \to B) \leftrightarrow (\neg B \to \neg A)$. This kind of generalization is the subject of a subsection below.

This justifies proof strategies we have already signalled above.

Proof Strategy: To prove a statement $A \to B$, we can aim instead for the equivalent $\neg B \to \neg A$: assume $\neg B$ and take $\neg A$ as our new goal. This is called "proving the contrapositive".

If we have posited both $A \to B$ and $\neg B$, then replacing the implication with the equivalent $\neg B \to \neg A$ and applying modus ponens allows us to conclude $\neg A$. The rule "From $A \to B$ and $\neg B$, conclude $\neg A$" is called *modus tollens*, and we have justified it.

We prove another theorem which justifies some additional proof strategies involving disjunction.

Theorem: $P \vee Q \leftrightarrow \neg P \to Q$

Corollary: $P \vee Q \leftrightarrow \neg Q \to P$. This follows from the theorem by equivalence of implications with their contrapositives and double negation.

Proof of Theorem: For part 1 of the proof, we assume $P \vee Q$ and deduce Goal 1: $\neg P \to Q$. The form of the posit suggests a proof by cases.

Case 1: We assume P. We prove the contrapositive of Goal 1: we assume $\neg Q$ and our goal is $\neg\neg P$. To prove $\neg\neg P$, we assume $\neg P$ and our goal is $\perp$, which is immediate as we have already posited P. This completes the proof of case 1.

Case 2: We assume Q. To prove the goal $\neg P \to Q$, we assume $\neg P$ and our new goal is Q. But we have already posited Q so we are done.

For part 2 of the proof, we assume $\neg P \to Q$ and deduce $P \vee Q$. We prove the goal by contradiction: we assume $\neg(P \vee Q)$ and take $\perp$ as our goal. We do this by proving P then proving $\neg P$. Our first goal is P, which we prove by contradiction: assume $\neg P$; by modus ponens Q follows, from which we can deduce $P \vee Q$, from which with our assumption $\neg(P \vee Q)$ we can deduce $\perp$, completing the proof of P by contradiction. Our second goal is $\neg P$: to prove this we assume P and take $\perp$ as our goal; from the assumption P we can deduce $P \vee Q$ from which with our assumption $\neg(P \vee Q)$ we can deduce $\perp$; this completes the proof of $\neg P$, which completes the proof by contradiction of $P \vee Q$.

Since the implications in both directions have been proved, the proof of the Theorem is complete.

The Theorem directly justifies the more general proof strategies for disjunction involving negative hypotheses given above.

Proof Strategy: To deduce the goal $A \vee B$, assume $\neg A$ and deduce B: this is valid because it is a proof of the equivalent implication $\neg A \rightarrow B$. Alternatively, assume $\neg B$ and deduce A: this is a proof of the equivalent $\neg B \rightarrow A$.

If we have posits $A \vee B$ and $\neg A$, we can draw the conclusion B, by converting $A \vee B$ to the equivalent $\neg A \rightarrow B$ and applying modus ponens.

Symmetrically, if we have posits $A \vee B$ and $\neg B$, we can deduce A.

A perhaps shocking result is that anything at all follows from $\perp$, and so from any contradiction.

Theorem: $\perp \rightarrow B$

Proof: Assume $\perp$, and our goal becomes B. We prove B by contradiction, that is, assume $\neg B$ and take $\perp$ as our new goal. The new goal is already met by our initial assumption, so the proof is complete.

Theorem: $A \wedge \neg A \rightarrow B$

Proof: Assume $A \wedge \neg A$, and take B as our new goal. From $A \wedge \neg A$, we deduce A and we deduce $\neg A$, and from these we deduce $\perp$. $\perp \rightarrow B$ is true by the previous theorem, and B follows by *modus ponens*.

The operation represented by $\neg$ is called *negation*.

2.7 Generality and the Rule of Substitution

A propositional letter P reveals nothing about the structure of the statement it denotes. This means that any argument that shows that certain hypotheses (possibly involving P) imply a certain conclusion A (possibly involving P) will remain valid if all occurrences of the propositional letter P in the entire context are replaced with any fixed statement B (which may be logically complex).

Denote the result of replacing P with B in A by $A[B/P]$. Extend this notation to sets Γ: $\Gamma[B/P] = \{A[B/P] \mid A \in \Gamma\}$.

The rule of substitution for propositional logic can then be stated as

If we can deduce A from a set of assumptions Γ, P is a propositional letter and B is any proposition (possibly complex), then we can deduce $A[B/P]$ from the assumptions $\Gamma[B/P]$.

Using the substitution notation, the strongest rules for the biconditional can be stated as

"from $A \leftrightarrow B$ and $C[B/P]$, deduce $C[A/P]$."

"from $A \leftrightarrow B$ and $C[A/P]$, deduce $C[B/P]$."

2.8 Note on Order of Operations

The statements "A and either B or C" and "Either A and B, or C" (which can formally be written $A \wedge (B \vee C)$ and $(A \wedge B) \vee C)$ do not have the same meaning. Making such grouping distinctions in English is awkward; in our notation we have the advantage of the mathematical device of parentheses.

To avoid having to write all parentheses in order to make the meaning of a statement clear, we stipulate that just as multiplication is carried out before addition when parentheses do not direct us to do otherwise, we carry out $\neg$ first, then $\wedge$, then $\vee$, then $\rightarrow$, then $\leftrightarrow$ or $\not\leftrightarrow$. When a list of operations at the same level are presented, we group to the right: $P \rightarrow Q \rightarrow R$ means $P \rightarrow (Q \rightarrow R)$. In fact, this only makes a difference for $\rightarrow$, as all the other basic operations are associative (including $\leftrightarrow$ and $\not\leftrightarrow$; check it out!).

There is a temptation to allow $A \leftrightarrow B \leftrightarrow C$ to mean $(A \leftrightarrow B) \wedge (B \leftrightarrow C)$ by forbidding the omission of parentheses in expressions $A \leftrightarrow (B \leftrightarrow C)$ and $(A \leftrightarrow B) \leftrightarrow C$. We resist this temptation.

2.9 Quantifiers

In this section, we go beyond propositional logic to what is variously called first-order logic, predicate logic, or the logic of quantifiers. In any event, as in the propositional logic section, we are not talking about a formal system, though we will introduce some formal notations: we are talking about kinds of statement which appear in informal mathematical argument in natural language.

We denote an arbitrary complex statement, presumably involving the variable x, by the notation $A[x]$. We do not write $A(x)$ because this is our notation for a unary predicate sentence in which A stands for some definite unary predicate: a sentence of the form $A(x)$ has the exact form of a predicate

being asserted of x while a sentence of the form $A[x]$ could be any sentence that presumably mentions x (so $x = x$ is of the form $A[x]$ but not of the form $A(x)$; a sentence like $\mathtt{Nat}(x)$ (meaning "x is a natural number") would be an example of the first form. A related notation is $A[t/x]$, the result of replacing the variable x in the proposition A with the term t (which may be a complex name rather than simply a variable). If we denote a formula $\mathcal{A}$ by the notation $A[x]$ then for any term t we use the notation $A[t]$ to represent $\mathcal{A}[t/x]$.[1]

The two kinds of statement we consider can be written "for all x, $A[x]$" (formulas with a *universal quantifier*) and "for some x, $A[x]$", which is also often written "there exists x such that $A[x]$" (which is why such formulas are said to have an *existential quantifier*). This language, although it is acceptable mathematical English, is already semi-formalized.

Formulas (or sentences) with universal and existential quantifiers can appear in a variety of forms. The statement "All men are mortal" can be analyzed as "for all x, if x is a man then x is mortal", and the statement "Some men are immortal" can be analyzed as "for some x, x is a man and x is immortal".

The formal notation for "for all x, $A[x]$ is $(\forall x.A[x])$ and for "for some x, $A[x]$" is $(\exists x.A[x])$. The parentheses in these notations are for us mandatory: this may seem eccentric but we think there are good reasons for it.

Iteration of the same quantifier can be abbreviated. We write $(\forall xy.A[x,y])$ instead of $(\forall x.(\forall y.A[x,y]))$, and similarly $(\exists xy.A[x,y])$ instead of $(\exists x.(\exists y.A[x,y]))$, and notations like $(\forall xyz.A[x,y,z])$ are defined similarly.

Quantifiers are sometimes (very often, in practice), *restricted* to some domain. Quantifiers restricted to a set have special notation: $(\forall x \in S.A[x])$ can be read "for all x in S, $A[x]$" and is equivalent to $(\forall x.x \in S \to A[x])$, while $(\exists x \in S.A[x])$ can be read "for some x in S, $A[x]$" and is equivalent to $(\exists x.x \in S \wedge A[x])$. The same quantifier restricted to the same set can be iterated, as in $(\forall xy \in S.A[x,y])$.

Further, restriction of a quantifier to a particular sort of object is not always explicitly indicated in the notation. If we know from the context that a variable n ranges over natural numbers, we can write $(\forall n.A[n])$ instead of $(\forall n \in \mathcal{N}.A[n])$, for example. In the section on typed theory of sets, all

[1] It should be noted that this is a subtle distinction I am drawing which is not universally made (the exact notation here is specific to these notes); it is quite common to write $P(x)$ for what I denote here as $P[x]$, and I have been known to write parentheses by mistake when teaching from this text.

variables will be equipped with an implicit type in this way.

We do not as a rule negate quantified sentences (or formulas) in natural language. Instead of saying "It is not the case that for all x, $A[x]$", we would say "For some x, $\neg A[x]$". Instead of saying "It is not the case that for some x, $A[x]$", we could say "For all x, $\neg A[x]$" (though English provides us with the construction "For no x, $A[x]$" for this case). "No men are mortal" means "For all x, if x is a man then x is not mortal". The logical transformations which can be carried out on negated quantified sentences are analogous to de Morgan's laws, and can be written formally

$$\neg(\forall x.A[x]) \leftrightarrow (\exists x.\neg A[x])$$

and

$$\neg(\exists x.A[x]) \leftrightarrow (\forall x.\neg A[x]).$$

Note that we are not licensing use of these equivalences as proof strategies before they are proved: as above with de Morgan's laws, these are introduced here to make a point about the rhetoric of mathematical English.

Here is a good place to say something formally about the distinction between the more general "formula" and the technical sense of "sentence" (I would really much rather say "sentence" for both, following the grammatical rather than the mathematical path). Any "sentence" in the grammatical sense of mathematical language is called a formula; the actual "sentences" in the mathematical sense are those in which a variable x only occurs in a context such as $(\forall x.A[x])$, $(\exists x.A[x])$ or even $\{x \mid A[x]\}$ or $\int_2^3 x^2\,dx$ (to get even more familiar) in which it is a dummy variable. The technical way of saying this is that a sentence is a formula in which all occurrences of any variable are *bound*.

2.10 Proving Quantified Statements and Using Quantified Hypotheses

To prove the goal "for all x, $A[x]$", introduce a new name a (not used before anywhere in the context): the new goal is to prove $A[a]$. Informally, if we can prove $A[a]$ without making any special assumptions about a, we have shown that $A[x]$ is true no matter what value the variable x takes on. The new name a is used only in this proof (its role is rather like that of an assumption in the proof of an implication).

To prove the goal "for some x, $A[x]$", find a specific name t (which may be complex) and prove $A[t]$. Notice here there may be all kinds of contextual knowledge about t and in fact that is expected. It's possible that several different such substitutions may be made in the course of a proof (in different cases a different witness may work to prove the existential statement).

If you have posited "for all x, $A[x]$", then you may further posit $A[t]$ for any name t, possibly complex. You may want to make several such substitutions in the course of a proof.

Using an existential statement is a bit trickier. If we have posited "for some x, $A[x]$", and we are aiming at a goal G, we may introduce a name w not mentioned anywhere in the context (and in particular not in G) and further posit $A[w]$: if G follows with the additional posit, it follows without it as well. What we are doing here is introducing a name for a witness to the existential hypothesis. Notice that this name is locally defined; it is not needed after the conclusion G is proved.

2.11 Equality and Uniqueness

For any term t, $t = t$ is an axiom which we may freely assert.

If we have posited $a = b$ and $A[a]$, we can further posit $A[b]$.

These are an adequate set of logical rules for equality.

To show that there is exactly one object x such that $A[x]$ (this is often written $(\exists! x.A[x])$), one needs to show two things: first, show $(\exists x.A[x])$ (there is at least one x). Then show that from the additional assumptions $A[a]$ and $A[b]$, where a and b are new variables not found elsewhere in the context, that we can prove $a = b$ (there is at most one x).

Proofs of uniqueness are often given in the form "Assume that $A[a]$, $A[b]$, and $a \neq b$: deduce a contradiction". This is equivalent to the proof strategy just given but the assumption $a \neq b$ is often in practice never used (one simply proves $a = b$) and so seems to be an unnecessary complication.

2.12 Dummy Variables and Substitution

The rules of the previous section make essential use of substitution. If we write the formula $A[x]$ of the previous section in the form $\mathcal{A}$, recall that the variants $A[a]$ and $A[t]$ mean $\mathcal{A}[a/x]$ and $\mathcal{A}[t/x]$: understanding these notations requires an understanding of substitution.

And there is something nontrivial to understand. Consider the sentence $(\exists x.x = a)$ (this is a sentence if a is a constant *name* rather than a variable). This is true for any a, so we might want to assert the not very profound theorem $(\forall y.(\exists x.x = y))$. Because this is a universal statement, we can drop the universal quantifier and replace y with anything to get another true statement: with c to get $(\exists x.x = c)$); with z to get $(\exists x.x = z)$). But if we naively replace y with x we get $(\exists x.x = x)$, which does not say what we want to say: we want to say that there is something which is equal to x, and instead we have said that there is something which is equal to *itself*.

The problem is that the x in $(\exists x.x = y)$ does not refer to any particular object (even if the variable x does refer to something in a larger context). x in this sentence is a "dummy variable". Since it is a dummy it can itself be replaced with any other variable: $(\exists w.w = y)$ means the same thing as $(\exists x.x = y)$, and replacing y with x in the former formula gives $(\exists w.w = x)$ which has the intended meaning.

Renaming dummy variables as needed to avoid collisions avoids these problems. We give a recursive definition of substitution which supports this idea. $T[t/x]$ is defined for T any term or formula, t any term, and x any variable. The only kind of term (noun phrase) that we have so far is variables: $y[t/x]$ is y if $y \neq x$ and t otherwise; $P(u)[t/x]$ is $P(u[t/x])$; $(u\,R\,v)[t/x]$ is $u[t/x]\,R\,v[t/x]$. So far we have defined substitution in such a way that it is simply replacement of the variable x by the term t. Where A is a formula which might contain x, $(\forall y.A)[t/x]$ is defined as $(\forall z.A[z/y][t/x])$, where z is the typographically first variable not occurring in $(\forall y.A)$, t or x. $(\exists y.A)[t/x]$ is defined as $(\exists z.A[z/y][t/x])$, where z is the typographically first variable not occurring in $(\exists y.A)$, t, or x. This applies to all constructions with bound variables, including term constructions: for example, once we introduce set notation, $\{y \mid A\}[t/x]$ will be defined as $\{z \mid A[z/y][t/x]\})$, where z is the typographically first variable not occurring in $\{y \mid A\}$, t, or x. The use of "typographically first" here is purely for precision: in fact our convention is that (for example) $(\forall x.A)$ is basically the same statement as $(\forall y.A[y/x])$ for any variable y not occurring in A (where our careful definition of substitution is used) so it does not matter which variable is used as long as the variable is new in the context.

It is worth noting that the same precautions need to be taken in carefully defining the notion of substitution for a propositional letter involved in the rule of substitution.

2.13 Are we doing formal logic yet?

One might think we are already doing formal logic. But from the strictest standpoint we are not. We have introduced formal notations extending our working mathematical language, but we are not yet considering terms, formulas and proofs *themselves* as mathematical objects and subjecting them to analysis (perhaps we are threatening to do so in the immediately preceding subsection). We will develop the tools we need to define terms and formulas as formal mathematical objects (actually, the tools we need to formally develop any mathematical object whatever) in the next section, and return to true formalization of logic (as opposed to development of formal notation) in the Logic section.

We have not given many examples: our feeling is that this material is so abstract that the best way to approach it is to use it when one has some content to reason about, which will happen in the next section. Reference back to our discussion of proof strategy here from actual proofs ahead of us is encouraged.

2.14 Exercises

Prove the following statements using the proof strategies above. Use only the highlighted proof strategies (not, for example, de Morgan's laws or the rules for negating quantifiers). You may use proof of an implication by proving the contrapositive, modus tollens and the generalized rules for proving disjunctions.

1. Prove the equivalence

$$A \to (B \to C) \leftrightarrow (A \wedge B) \to C$$

2. Prove the equivalence

$$\neg(A \to B) \leftrightarrow (A \wedge \neg B)$$

3. Prove

$$((P \to Q) \wedge (Q \to R)) \to (P \to R)$$

4. Prove

$$\neg(\forall x.P[x]) \leftrightarrow (\exists x.\neg P[x])$$

5. Prove

$$((\exists x.P[x]) \wedge (\forall uv.P[u] \to Q[v])) \to (\forall z.Q[z])$$

6. Prove de Morgan's laws.

We give some solutions.

1: Our goal is

$$A \to (B \to C) \leftrightarrow (A \wedge B) \to C.$$

Goal 1:

$$A \to (B \to C) \to (A \wedge B) \to C$$

.

Argument for Goal 1: Assume $A \to (B \to C)$. Our new goal is $(A \wedge B) \to C$. To prove this implication we further assume $A \wedge B$, and our new goal is C. Since we have posited $A \wedge B$, we may deduce both A and B separately. Since we have posited A and $A \to (B \to C)$ we may deduce $B \to C$ by modus ponens. Since we have posited B and $B \to C$, we may deduce C, which is our goal, completing the proof of Goal 1.

Goal 2:

$$(A \wedge B) \to C \to A \to (B \to C)$$

Argument for Goal 2: Assume $(A \wedge B) \to C$. Our new goal is $A \to (B \to C)$. To deduce this implication, we assume A and our new goal is $B \to C$. To deduce this implication, we assume B and our new goal is C. Since we have posited both A and B we may deduce $A \wedge B$. Since we have posited $A \wedge B$ and $(A \wedge B) \to C$, we may deduce C by modus ponens, which is our goal, completing the proof of Goal 2.

Conclusion: Since the implications in both directions have been proved, the biconditional main goal has been proved.

3: Our goal is

$$\neg(\forall x.P[x]) \leftrightarrow (\exists x.\neg P[x]).$$

Since this is a biconditional, the proof involves proving two subgoals.

Goal 1:

$$\neg(\forall x.P[x]) \rightarrow (\exists x.\neg P[x])$$

Argument for Goal 1: Assume $\neg(\forall x.P[x])$. Our new goal is $(\exists x.\neg P[x])$. We would like to prove this by exhibiting a witness, but we have no information about any specific objects, so our only hope is to start a proof by contradiction. We assume $\neg(\exists x.\neg P[x])$ and our new goal is $\bot$. We note that deducing $(\forall x.P[x])$ as a goal would allow us to deduce $\bot$ (this is one of the main ways to use a negative hypothesis). To prove this goal, introduce an arbitrary object a and our new goal is $P[a]$. Since there is no other evident way to proceed, we start a new proof by contradiction: assume $\neg P[a]$ and our new goal is $\bot$. Since we have posited $\neg P[a]$, we may deduce $(\exists x.\neg P[x])$. This allows us to deduce $\bot$, since we have already posited the negation of this statement. This supplies what is needed for each goal in turn back to Goal 1, which is thus proved.

Goal 2:

$$(\exists x.\neg P[x]) \rightarrow \neg(\forall x.P[x])$$

Argument for Goal 2: We assume $(\exists x.\neg P[x])$. Our new goal is $\neg(\forall x.P[x])$. To deduce this goal, we assume $(\forall x.P[x])$ and our new goal is $\bot$. Our existential hypothesis $(\exists x.\neg P[x])$ allows us to introduce a new object a such that $\neg P[a]$ holds. But our universal hypothesis $(\forall x.P[x])$ allows us to deduce $P[a]$ as well, so we can deduce $\bot$, completing the proof of Goal 2.

Conclusion: Since both implications involved in the biconditional main goal have been proved, we have proved the main goal.

3 Typed theory of sets

In this section we introduce a theory of sets, but not the usual one quite yet. We choose to introduce a typed theory of sets, which might carelessly be attributed to Russell, though historically this is not quite correct.

3.1 Types in General

Mathematical objects come in sorts or kinds (the usual word is "type"). We seldom make any statement about all mathematical objects whatsoever: we are more likely to be talking about all natural numbers, or all real numbers. In much of this section, every variable we introduce will have a type, and a quantifier over that variable will be implicitly restricted to that type.

3.2 Typed Theory of Sets

We introduce a typed theory of sets in this section, loosely based on the historical type theory of Bertrand Russell. This theory is sufficiently general to allow the construction of all objects considered in classical mathematics. We will demonstrate this by carrying out some constructions of familiar mathematical systems. An advantage of using this type theory is that the constructions we introduce will not be the same as those you might have seen in other contexts, which will encourage careful attention to the constructions and proofs, which furthers other parts of our implicit agenda. Later we will introduce a more familiar kind of set theory.

Suppose we are given some sort of mathematical object (natural numbers, for example). Then it is natural to consider collections of natural numbers as another sort of object. Similarly, when we are given real numbers as a sort of object, our attention may pass to collections of real numbers as another sort of object.

Our approach is an abstraction from this. We introduce a sort of mathematical object which we will call *individuals* about which we initially assume nothing whatsoever (we will add an axiom asserting that there are infinitely many individuals when we see how to say this). We also call the sort of individuals *type 0*. We then define *type 1* as the sort of collections of individuals, *type 2* as the sort of collections of type 1 objects, and so forth.

No essential role is played here by natural numbers: we could call type 0 ι and for any type τ let τ^+ be the sort of collections of type τ objects, and then

the types 0,1,2... would be denoted $\iota, \iota^+, \iota^{++}, \ldots$ in which we can see that no reference to natural numbers is involved. This paragraph is an answer in advance to an objection raised by philosophers: later we will define the natural number 3 (for example) in type theory: we have not assumed that we already understand what 3 is by using "3" as a formal name for the type ι^{+++}.

Every variable x comes equipped with a type. We may write $x^{\mathbf{3}}$ for a type 3 variable (type superscripts will be boldface when they do appear so as not to be confused with exponents or other numerical superscripts), but we will not always do this. Atomic formulas of our language are of the form $x = y$, in which the variables x and y must be of the same type, and $x \in y$ in which the type of y must be the successor of the type of x.

Just for fun we give a formal description of the grammatical requirements for formulas which does not use numerals (in fact, amusingly, it does not even mention types!). Please note that we will not actually *use* the notation outlined in this paragraph: the point is that the notation we actually use could be taken as an abbreviation for this notation, which makes the point firmly that we are not actually assuming that we know anything about natural numbers yet when we use numerals as type superscripts. We use a more long-winded notation for variables. We make the following stipulations: $\mathbf{x}$ is an individual variable; if y is an individual variable, so is y'; these two rules ensure that we have infinitely many distinct individual (type 0, but we aren't mentioning numerals) variables. Now we define variables in general: an individual variable is a variable; if y is a variable, y^+ is a variable (one type higher, but we are not mentioning numerals). Now we define grammatical atomic formulas. If x is an individual variable and y is a variable, then $x = y$ is an atomic formula iff y is an individual variable. If x is an individual variable, then $x \in y$ is an atomic formula iff y is of the form z^+ where z is an individual variable. For any variables x and y, $x^+ = y^+$ is an atomic formula iff $x = y$ is an atomic formula and $x^+ \in y^+$ is an atomic formula iff $x \in y$ is an atomic formula. We do not write any atomic formula which we cannot show to be grammatical using these rules. The variable consisting of x followed by m primes and n plusses might more conveniently be written $x^{\mathbf{n}}_m$, but in some formal sense it does not have to be: there is no essential reference to numerals here. The rest of the formal definition of formulas: if ϕ is an atomic formula, it is a formula; if ϕ and ψ are formulas and x is a variable, so are (ϕ), $\neg\phi$, $\phi \wedge \psi$, $\phi \vee \psi$, $\phi \rightarrow \psi$, $\phi \leftrightarrow \psi$, $(\forall x.\psi)$, $(\exists x.\psi)$ [interpreting formulas with propositional connectives is made more complicated

by order of operations, but the details are best left to a computer parser!].

Our theory has axioms. The inhabitants of every type other than 0 are sets. We believe that sets are equal iff they have exactly the same elements. This could be expressed as follows:

*__Strong axiom of extensionality:__

$$(\forall x.(\forall y.x = y \leftrightarrow (\forall z.z \in x \leftrightarrow z \in y))),$$

for every assignment of types to x, y, z that makes sense.

*__Proof Strategy:__ If A and B are sets, to prove $A = B$, introduce a new variable a, assume $a \in A$, and deduce $a \in B$, and then introduce a new variable b, assume $b \in B$, and deduce $b \in A$. This strategy simply unfolds the logical structure of the axiom of extensionality.

This axiom says that objects of any positive type are equal iff they have the same elements. This is the natural criterion for equality between sets.

Notice that we did not write

$$(\forall x^{\mathbf{n+1}}.(\forall y^{\mathbf{n+1}}.x^{\mathbf{n+1}} = y^{\mathbf{n+1}} \leftrightarrow (\forall z^{\mathbf{n}}.z^{\mathbf{n}} \in x^{\mathbf{n+1}} \leftrightarrow z^{\mathbf{n}} \in y^{\mathbf{n+1}}))).$$

This would be very cumbersome, and it is not necessary: it is clear from the form of the sentence (it really is a sentence!) that x and y have to have the same type (because $x = y$ appears) and z has to be one level lower in type (because $z \in x$ appears). One does need to be careful when taking this implicit approach to typing to make sure that everything one says *can* be expressed in the more cumbersome notation: more on this anon.

Notice that we starred the strong axiom of extensionality; this is because it is not the axiom we actually adopt. We take the more subtle view that in the real world not all objects are sets, so we might want to allow many non-sets with no elements (it is reasonable to suppose that anything with an element *is* a set). Among the objects with no elements, we designate a particular object $\emptyset$ as the empty *set*.

This does mean that we are making our picture of the hierarchy of types less precise: type $n + 1$ is inhabited by collections of type n objects and also possibly by other junk of an unspecified nature.

Primitive notion: There is a designated object $\emptyset^{\mathbf{n+1}}$ for each positive type $n + 1$ called the *empty set* of type $n + 1$. We do not always write the type index.

Axiom of the empty set: $(\forall x.x \notin \emptyset)$, for all assignments of a type to x and $\emptyset$ which make sense.

Definition: We say that an object x (in a positive type) is a *set* iff $x = \emptyset \vee (\exists y.y \in x)$. We write $\mathtt{set}(x)$ to abbreviate "x is a set" in formulas. We say that objects which are not sets are *atoms* or *urelements*.

Axiom of extensionality:

$$(\forall xy.\mathtt{set}(x) \wedge \mathtt{set}(y) \rightarrow x = y \leftrightarrow (\forall z.z \in x \leftrightarrow z \in y)),$$

for any assignment of types to variables that makes sense.

Proof Strategy: If A and B are sets, to prove $A = B$, introduce a new variable a, assume $a \in A$, and deduce $a \in B$, and then introduce a new variable b, assume $b \in B$, and deduce $b \in A$. This strategy simply unfolds the logical structure of the axiom of extensionality.

We have already stated a philosophical reason for using a weaker form of the axiom of extensionality, though it may not be clear that this is applicable to the context of type theory (one might reasonably suppose that non-sets are all of type 0); we will see mathematical reasons for adopting the weaker form of extensionality in the course of our development (and we will also see mathematical advantages of strong extensionality).

We have said when sets are equal. Now we ask what sets there are. The natural idea is that any property of type n objects should determine a set of type $n + 1$, and this is what we will say:

Axiom of comprehension: For any formula $A[x]$ in which the variable y (of type one higher than x) does not appear,

$$(\exists y.(\forall x.x \in y \leftrightarrow A[x])).$$

This says that for any formula $A[x]$ expressing a property of an object x (of some type n), there is an object y of type $n + 1$ such that the elements of y are exactly the objects x such that $A[x]$.

The axiom of extensionality tells us that there is only one such object y which is a set (there may be many such objects y if $A[x]$ is not true for any x, but only one of them ($\emptyset$) will be a set). This suggests a definition:

Set builder notation: For any formula $A[x]$, define $\{x \mid A[x]\}$ as the unique *set* of all x such that $A[x]$: this exists by Comprehension and is uniquely determined by Extensionality. If x is of type n, then $\{x \mid A[x]\}$ is of type $n+1$.

Proof Strategy: To use a posit or deduce a goal of the form $t \in \{x \mid A[x]\}$, replace the posit or goal with the equivalent $A[t]$.

In our numeral free notation we indicate the grammar requirements for set abstracts: if x is a variable and ϕ is a formula, $\{x \mid \phi\}$ can replace any occurrence of x^+ in a formula and it will still be a formula.

There are two other axioms in our system, the Axiom of Infinity and the Axiom of Choice, but some formal development should be carried out before we introduce them.

3.3 Russell's Paradox?

At this point an objection might interpose itself. Consider the following argument.

For any set x, obviously either x is an element of itself or x is not an element of itself. Form the set R whose elements are exactly those sets which are not elements of themselves: $R = \{x \mid x \notin x\}$. Now we ask, is R an element of itself? For any x, $x \in R \leftrightarrow x \notin x$, so in particular $R \in R \leftrightarrow R \notin R$. This is a contradiction!

This argument, known as *Russell's paradox*, was a considerable embarrassment to early efforts to formalize mathematics on the very abstract level to which we are ascending here.

Fortunately, it is completely irrelevant to our work here. This argument does not work in our system, on a purely formal level, because $x \in x$ is not a legal formula in the language of our type theory, so it does not define a property of sets allowing the introduction of a set by Comprehension! On a less formal level, attending to the meaning of notations rather than their formal structure, we have not introduced the kind of sweeping notion of set presupposed in the argument for Russell's paradox: for any particular sort of object τ (such as type n) we have introduced the new sort of object "set of τ's" or "τ^+" (which we call type $n+1$ in the particular case where τ is type n). The supposition in Russell's paradox is that we have a type of sets which contains all sets of objects of that same type. Ordinary mathematical constructions do not lead us to a situation where we need such a type.

If we had a universal sort $\mathtt{o}$ containing *all objects* it might seem that $\mathtt{o}^+$ would contain all sets of anything whatsoever (including sets of type $\mathtt{o}^+$ sets, which would presumably also be of the universal type $\mathtt{o}$). The argument for Russell's paradox shows that there cannot be such a type if the Axiom of Comprehension is to apply: either there cannot be a universal type $\mathtt{o}$ or the type $\mathtt{o}^+$ cannot contain all definable subcollections of $\mathtt{o}$. We will introduce untyped set theories with restrictions on comprehension below.

It is important to notice on a philosophical level that care in the introduction of the idea of a set has completely avoided the paradox: there is no embarrassment for our typed notion of set, and our typed notion of set is true to what we actually do in mathematics. Russell's paradox was a serious problem for an initial insufficiently careful development of the foundation of mathematics; it is not actually a problem for the foundations of mathematics as such, because the typed notion of set is all that actually occurs in mathematics in practice (in spite of the fact that the system of set theory which is customarily used is formally untyped: we shall meet this system in section 4 and see that its restrictions on comprehension can be naturally motivated in terms of types).

Notice that if x and y are terms of different types, $x = y$ is not a formula at all. This does not mean that we say that x and y are distinct: it means that we do not entertain the question as to whether objects of different types are identical or distinct (for now; we will have occasion to think about this later). Similarly, if the type of y is not the successor of the type of x (for example, if x and y are of the same type) we do not say $x \in y$ (it is ungrammatical, not false). We do not ask whether $x \in x$; we do not say that it is false (or true) (for now).

3.4 Simple Ideas of Set Theory

In this section we develop some familiar ideas of set theory.

We first develop the familiar list notation for finite sets. Here are the standard notations for one and two element sets.

List notation for sets: $\{x\}$ is defined as $\{y \mid y = x\}$. $\{x, y\}$ is defined as $\{z \mid z = x \vee z = y\}$.

It is convenient to define Boolean union and intersection of sets before giving the general definition of list notation.

Boolean union and intersection: If x and y are sets, define $x \cup y$ as

$$\{z \mid z \in x \vee z \in y\}$$

and $x \cap y$ as

$$\{z \mid z \in x \wedge z \in y\}.$$

Notice that though we may informally think of $x \cup y$ as "x and y", it is actually the case that $x \cup y$ is associated with the logical connective $\vee$ and it is $x \cap y$ that is associated with $\wedge$ in a logical sense.

We also define a^c (the complement of a) as $\{x \mid x \notin a\}$ and $a - b$ (the set difference of a and b) as $a \cap b^c$.

recursive definition of list notation: $\{x_1, x_2, \ldots, x_n\}$ is defined as $\{x_1\} \cup \{x_2, \ldots, x_n\}$. Notice that the definition of list notation for n items presupposes the definition of list notation for $n-1$ items: since we have a definition of list notation for 1 and 2 items we have a basis for this recursion.

Note that all elements of a set defined by listing must be of the same type, just as with any set.

There is one more very special case of finite sets which needs special attention.

null set: We have introduced $\emptyset^{\mathbf{n+1}}$ as a primitive notion because we adopted the weak axiom of extensionality.

If we assumed strong extensionality, we could define $\emptyset^{\mathbf{n+1}}$ as

$$\{x^{\mathbf{n}} \mid x^{\mathbf{n}} \neq x^{\mathbf{n}}\}$$

(in any event this set abstract is equal to $\emptyset^{\mathbf{n+1}}$!). Notice that $\emptyset^{\mathbf{n+1}}$ has no elements, and it is by Extensionality (either form) the only set (of type $n+1$) with no elements. In this definition we have used type superscripts, though hereinafter we will write just $\emptyset$: this is to emphasize that $\emptyset$ is defined in each positive type and we do not say that the empty sets in different types are the same (or that they are different). Notice that although $x \in x$ is not grammatical, $\emptyset \in \emptyset$ *is* grammatical (and false!). It is not an instance of the ungrammatical form $x \in x$ because the apparent identity of the two occurrences of $\emptyset$ is a kind of pun. The pun can be dispelled by writing $\emptyset^{\mathbf{n+1}} \in \emptyset^{\mathbf{n+2}}$ explicitly.

universe: We define V as $\{x \mid x = x\}$. This is the universal set. The universal set in type $n+1$ is the set of all type n objects. $V \in V$ is grammatical and true – but the two occurrences of V have different reference (this can be written $V^{\mathbf{n+1}} \in V^{\mathbf{n+2}}$ for clarification).

Of course we assume that the universal set is not finite, but we do not know how to say this yet.

The combination of the empty set and list notation allows us to write things like $\{\emptyset, \{\emptyset\}\}$, but not things like $\{x, \{x\}\}$: the former expression is another pun, with empty sets of different types appearing, and the latter expression is ungrammatical, because it is impossible to make a consistent type assignment to x. An expression like this can make sense in an untyped set theory (and in fact in the usual set theory the first expression here is the most popular way to define the numeral 2, as we will explain later).

Set builder notation can be generalized.

Generalized set builder notation: If we have a complex term $t[x_1, \ldots, x_n]$ containing only the indicated variables, we define $\{t[x_1, \ldots, x_n] \mid A\}$ as $\{y \mid (\exists x_1 \ldots x_n.y = t[x_1, \ldots, x_n] \wedge A)\}$ (where y is a new variable). We do know that this kind of very abstract definition is not really intelligible in practice except by backward reference from examples, and we will provide these!

Examples: $\{\{x\} \mid x = x\}$ means, by the above convention, $\{z \mid (\exists z.z = \{x\} \wedge z = z)\}$. It is straightforward to establish that this is the set of all sets with exactly one element, and we will see below that we will call this the natural number 1. The notation $\{\{x, y\} \mid x \neq y\}$ expands out to $\{z \mid (\exists xy.z = \{x, y\} \wedge x \neq y)\}$: this can be seen to be the set of all sets with exactly two elements, and we will identify this set with the natural number 2 below.

We define some familiar relations on sets.

subset, superset: We define $A \subseteq B$ as

$$\mathtt{set}(A) \wedge \mathtt{set}(B) \wedge (\forall x.x \in A \rightarrow x \in B).$$

We define $A \supseteq B$ as $B \subseteq A$.

Theorem: For any set A, $A \subseteq A$.

Theorem: For any sets A, B, $A \subseteq B \wedge B \subseteq A \rightarrow A = B$.

Theorem: For any sets A, B, C, if $A \subseteq B$ and $B \subseteq C$ then $A \subseteq C$.

Observation: The theorems we have just noted will shortly be seen to establish that the subset relation is a "partial order".

Proof Strategy: To show that $A \subseteq B$, where A and B are known to be sets, introduce an arbitrary object x and assume $x \in A$: show that it follows that $x \in B$.

If one has a hypothesis or previously proved statement $A \subseteq B$ and a statement $t \in A$, deduce $t \in B$.

Notice that the proof strategy given above for proving $A = B$ is equivalent to first proving $A \subseteq B$, then proving $B \subseteq A$.

The notions of element and subset can be confused, particularly because we have a bad habit of saying things like "A is in B" or "A is contained in B" for $A \in B$ and for $A \subseteq B$. It is useful to observe that elements are not "parts" of sets. The relation of part to whole is transitive: if A is a part of B and B is a part of C, then A is a part of C. The membership "relation" is not transitive in a quite severe sense: if $A \in B$ and $B \in C$, then $A \in C$ is not even meaningful in our type theory! [In the untyped set theories discussed in section 4, membership is in a quite normal sense not transitive.] But the subset relation is transitive: if $A \subseteq B$ and $B \subseteq C$, then any element of A is also an element of B, and so is in turn an element of C, so $A \subseteq C$. If a set can be said to have parts, they will be its subsets, and its one-element sets $\{a\}$ for $a \in A$ can be said to be its atomic parts.

We give a general format for introducing operations, and then introduce an important operation.

Definable Operations: For any formula $\phi[x, y]$ with the property that

$$(\forall xyz.\phi[x, y] \wedge \phi[x, z] \rightarrow y = z)$$

we define $F_\phi(x)$ or $F_\phi{}'x$ as the unique y (if there is one) such that $\phi[x, y]$. Note that we will not always explicitly give a formula ϕ defining an operation, but it should always be clear that such a formula could be given. Note also that there might be a type differential between x and $F_\phi(x)$ depending on the structure of the formula $\phi[x, y]$.

For any such definable operation $F(x)$, we define F"x for any set x as $\{F(u) \mid u \in x\}$: F"x is called the (elementwise) *image* of x under the operation F.

Power Set: For any set A, we define $\mathcal{P}(A)$ as $\{B \mid B \subseteq A\}$. The power set of A is the set of all subsets of A. Notice that $\mathcal{P}(V^{\mathbf{n}})$ is the collection of all sets of type $n+1$, and is not necessarily the universe $V^{\mathbf{n+1}}$, which might also contain some atoms.

Observation: $\mathcal{P}$ is F_ϕ where $\phi[x, y]$ is the formula $(\forall z.z \in y \leftrightarrow z \subseteq x)$ (or just $y = \{z \mid z \subseteq x\}$).

It is **very important** to notice that $\mathcal{P}(x)$ is one type higher than x, and similarly that $\{x\}$ is one type higher than x.

In the usual untyped set theory, the natural numbers are usually defined using a clever scheme due to John von Neumann.

***Definition:** 0 is defined as $\emptyset$. 1 is defined as $\{0\}$. 2 is defined as $\{0, 1\}$. 3 is defined as $\{0, 1, 2\}$. In general, $n + 1$ is defined as $n \cup \{n\}$.

The star on this "definition" indicates that we do not use it here. The problem is that this definition makes no sense in our typed language. Notice that there is no consistent way to assign a type to n in "$n \cup \{n\}$". In section 4 on untyped set theory, we will be able to use this definition and we will see that it generalizes to an incredibly slick definition of ordinal number.

NOTE: elementary theorems should follow (proof examples!). Definition and discussion of boolean algebras?

3.5 Finite Number; the Axiom of Infinity; Ordered Pairs

The motivation of our definition of natural number in type theory is the following

Circular Definition: The natural number n is the set of all sets with n elements.

Of course this will not be acceptable as a formal definition: we spend the rest of the section showing how we can implement it using a series of formally valid definitions.

It is amusing to observe that the von Neumann definition above can also be motivated using an even worse

*__Circular Definition:__ The natural number n is the set of all natural numbers less than n.

This is starred to indicate that we are not at this point using it at all!

Definition: We define 0 as $\{\emptyset\}$.

Note that we have thus defined 0 as the set of all sets with zero (no) elements.

Definition: For any set A, define $A+1$ as $\{x \cup \{y\} \mid x \in A \wedge y \notin x\}$. $A+1$ is the collection of all sets obtained by adjoining a single new element to an element of A.

Definition: We define 1 as $0+1$. (Observe that 1 is the set of all one-element sets (singletons).) We define 2 as 1+1, 3 as 2+1, and so forth (and observe that 2 is the set of all sets with exactly two elements, 3 is the set of all sets with exactly three elements, and so forth).

Unfortunately, "and so forth" is a warning that a careful formal examination is needed at this point!

Definition: We call a set I an *inductive set* if $0 \in I$ and $(\forall A.A \in I \rightarrow A+1 \in I)$. We define $\mathcal{I}$ as the set of all inductive sets.

At this point it is useful to define the unions and intersections of not necessarily finite collections of sets.

Definition: For any set A, we define $\bigcup A$ as

$$\{x \mid (\exists a \in A.x \in a)\}$$

and $\bigcap A$ as

$$\{x \mid (\forall a \in A.x \in a)\}.$$

(Notice that $x \cup y = \bigcup\{x, y\}$ and $x \cap y = \bigcap\{x, y\}$.)

Observation: Notice that $\bigcup A$ and $\bigcap A$ are of type $n+1$ if A is of type $n+2$.

Definition: We define $\mathbb{N}$, the set of all natural numbers, as $\bigcap \mathcal{I}$, the intersection of all inductive sets.

We saw above that 0 has been successfully defined as the set of all zero element sets, 1 as the set of all one-element sets, 2 as the set of all two-element sets and so forth (whenever and so forth, etc, ... or similar devices appear in mathematical talk, it is a signal that there is something the author hopes you will see so that he or she does not have to explain it!) So we can believe for each of the familiar natural numbers (as far as we care to count) that we have implemented it as a set. If I is an inductive set, we can see that (the set implementing) 0 is in I by the definition of "inductive". If the set implementing the familiar natural number n is in I, then (by definition of "inductive") the set implementing the familiar natural number $n+1$ will be in I. So by the principle of mathematical induction, sets implementing each of the familiar natural numbers are in I. But I was *any* inductive set, so for each familiar natural number n, the set implementing n is in the intersection of all inductive sets, that is in $\mathbb{N}$ as we have defined it. This is why we call inductive sets "inductive", by the way. How can we be sure that there aren't some other unintended elements of $\mathbb{N}$? The best argument we can give is this: if there is a collection containing exactly the implementations of the familiar natural numbers, we observe that 0 is certainly in it and $n+1$ must be in it if n is in it. So this collection is inductive, so any element of $\mathbb{N}$, the intersection of all inductive sets, must belong to this set too, and so must be one of the familiar natural numbers. We will see later that there are models of type theory (and of untyped set theory) in which there *are* "unintended" elements of $\mathbb{N}$. In such models the collection of familiar natural numbers must fail to be a set. How can this happen when each type $k+1$ is supposed to be the collection of *all* sets of type k objects? Notice that the axiom of comprehension only forces us to implement the subcollections of type k which are definable using a formula of our language as type $k+1$ objects. So if there are "unintended" natural numbers we will find that no formula of our language will pick out just the familiar natural numbers. If we insist that each type $k+1$ contain *all* collections of type k objects, it will follow that we have defined the set of natural numbers correctly.

Definition: We define $\mathbb{F}$, the set of all finite sets, as $\bigcup \mathbb{N}$. A set which is not finite (not an element of $\mathbb{F}$) is said to be *infinite*.

Since we have defined each natural number n as the set of all sets with n elements, this is the correct definition of finite set (a finite set is a set which has n elements for some natural number n, so exactly a set which belongs to n for some $n \in \mathbb{N}$).

Now we can state a promised axiom.

Axiom of Infinity: $V \notin \mathbb{F}$

This says exactly that the universe is infinite.

In all of this, we have not issued the usual warnings about types. We summarize them here. For $A+1$ to be defined, a set must be of at least type 2. $A+1$ is of the same type as A. Similarly, 0 is of type at least 2 (and there is a formally distinct $0^{\mathbf{n+2}}$ for each n). Any inductive set must be of at least type 3 and the set of all inductive sets $\mathcal{I}$ is of at least type 4. $\mathbb{N}$ is then of type at least 3 (it being the minimal inductive set) and there is actually a $\mathbb{N}^{\mathbf{n+3}}$ in each type $n+3$. An amusing pun which you may check is $0 \in 1$. The Axiom of Infinity, like the two earlier axioms, says something about each type: the universal set over each type is infinite (it could be written more precisely as $V^{\mathbf{n+1}} \notin \mathbb{F}^{\mathbf{n+2}}$).

We state basic properties of the natural numbers. These are Peano's axioms for arithmetic in their original form. The theory with these axioms (which makes essential use of sets of natural numbers in its formulation) is called *second-order Peano arithmetic*.

1. $0 \in \mathbb{N}$
2. For each $n \in \mathbb{N}$, $n+1 \in \mathbb{N}$.
3. For all $n \in \mathbb{N}$, $n+1 \neq 0$
4. For all $m, n \in \mathbb{N}$, $m+1 = n+1 \rightarrow m = n$.
5. For any set $I \subseteq \mathbb{N}$ such that $0 \in I$ and for all $n \in I$, $n+1 \in I$, all natural numbers belong to I (the principle of mathematical induction).

All of these are obvious from the definition of $\mathbb{N}$ except axiom 4. It is axiom 4 that hinges on the adoption of the Axiom of Infinity.

The principle of mathematical induction (axiom 5) can be presented as another

Proof Strategy: To deduce a goal

$$(\forall n \in \mathbb{N}.\phi[n]),$$

define A as the set $\{n \in \mathbb{N} \mid \phi[n]\}$ and deduce the following goals:

Basis step: $0 \in A$

Induction step: The goal is $(\forall k \in \mathbb{N} \mid k \in A \rightarrow k+1 \in A)$: to prove this, let k be an arbitary natural number, assume $k \in A$ (equivalently $\phi[k]$) (called the *inductive hypothesis*) and deduce the new goal $k+1 \in A$ (equivalently $\phi[k+1]$).

We prove some theorems about natural numbers.

Theorem (not using Infinity): For any natural number n, if $x \in n+1$ and $y \in x$, then $x - \{y\} \in n$. [an equivalent form is "if $x \cup \{y\} \in n+1$ then $x - \{y\} \in n$"]

Proof: Let $A = \{n \in \mathbb{N} \mid (\forall xy.x \in n+1 \wedge y \in x \rightarrow x - \{y\} \in n)\}$, i.e., the set of all n for which the theorem is true. Our strategy is to show that the set A is inductive. This is sufficient because an inductive set will contain all natural numbers.

First Goal: $0 \in A$

Proof of First Goal: The goal is equivalent to the assertion that if $x \in 0+1$ and $y \in x$, then $x - \{y\} \in 0$. We suppose that $x \in 0+1 = 1$ and $y \in x$: this implies immediately that $x = \{y\}$, whence we can draw the conclusion $x - \{y\} = \{y\} - \{y\} = \emptyset \in 0$, and $x - \{y\} \in 0$ is our first goal.

Second Goal: $(\forall k \in A.k+1 \in A)$

Proof of Second Goal: Let k be an element of A. Assume that $k \in A$: this means that for any $x \in k+1$ and $y \in x$ we have $x - \{y\} \in k$ (this is the inductive hypothesis). Our goal is $k+1 \in A$: we need to show that if $u \in (k+1)+1$ and $v \in u$ we have $u - \{v\} \in k+1$. So we assume $u \in (k+1)+1$ and $v \in u$: our new goal is $u - \{v\} \in k+1$. We know because $u \in (k+1)+1$ that there are $p \in k+1$ and $q \notin p$ such that $p \cup \{q\} = u$. We consider two cases: either $v = q$ or $v \neq q$. If $v = q$ then $u - \{v\} = (p \cup \{q\}) - \{q\} = p$ (because $q \notin p$)

and we have $p \in k+1$ so we have $u-\{v\} \in k+1$. In the case where $v \neq q$, we have $v \in p$, so $p - \{v\} \in k$ by the inductive hypothesis, and $u - \{v\} = (p - \{v\}) \cup \{q\} \in k + 1$ because $p - \{v\} \in k$ and $q \notin p - \{v\}$. In either case we have the desired goal so we are done.

Theorem (not using Infinity): If n is a natural number and $x, y \in n$ and $x \subseteq y$ then $x = y$.

Proof: Let A be the set of natural numbers for which the theorem is true: $A = \{n \in \mathbb{N} \mid (\forall xy.x \in n \wedge y \in n \wedge x \subseteq y \rightarrow x = y)\}$. Our strategy is to show that A is inductive.

First Goal: $0 \in A$

Proof of First Goal: What we need to prove is that if $x \in 0$ and $y \in 0$ and $x \subseteq y$ then $x = y$. Assume that $x \in 0$ and $y \in 0$ and $x \subseteq y$. It follows that $x = \emptyset$ and $y = \emptyset$, so $x = y$. This completes the proof. Note that the hypothesis $x \subseteq y$ did not need to be used.

Second Goal: $(\forall k \in A.k + 1 \in A)$

Proof of Second Goal: Assume $k \in A$. This means that for all $x, y \in k$, if $x \subseteq y$ then $x = y$. This is called the inductive hypothesis.

Our goal is $k + 1 \in A$. This means that for all $u, v \in k + 1$, if $u \subseteq v$ then $u = v$. Suppose that $u \in k + 1$, $v \in k + 1$, and $u \subseteq v$. Our goal is now $u = v$. Because $u \in k + 1$, there are a and b such that $u = a \cup \{b\}$, $a \in k$, and $b \notin a$. Because $u \subseteq v$ we have $a = u - \{b\} \subseteq v - \{b\}$. $a \in k$ has been assumed and $v - \{b\} \in k$ by the previous theorem ($b \in v$ because $u \subseteq v$), so $a = v - \{b\}$ by inductive hypothesis, so $u = a \cup \{b\} = (v - \{b\}) \cup \{b\} = v$.

Theorem (not using Infinity): If there is a natural number n such that $V \in n$, we have $n = \{V\}$, $n + 1 = \emptyset \in \mathbb{N}$, and $n + 1 = \emptyset + 1$, though $n \neq \emptyset$, a counterexample to Axiom 4.

Proof: If $V \in n \in \mathbb{N}$, then for any $x \in n$ we clearly have $x \subseteq V$ whence $x = V$ by the previous theorem, so $n = \{V\}$. That $\{V\} + 1 = \emptyset$ is obvious from the definition of successor (we cannot add a new element

to V). It then clearly follows that $\emptyset$ is a natural number. $\emptyset{+}1 = \emptyset$ is also obvious from the definition of successor, so we get the counterexample to Axiom 4.

Theorem (using Infinity): $(\forall mn \in \mathbb{N}.m + 1 = n + 1 \rightarrow m = n)$.

Proof: Suppose that m and n are natural numbers and $m + 1 = n + 1$.

We prove that $m = n$ by showing that they have the same elements.

Let $a \in m$ be chosen arbitrarily: our aim is to show $a \in n$.

Choose $x \notin a$ (that there is such an x follows from the Axiom of Infinity, which tells us that the finite set a (finite because it belongs to a natural number) cannot be V). $a \cup \{x\} \in m + 1$. It follows that $a \cup \{x\} \in n + 1$, since by hypothesis $m + 1 = n + 1$. It then follows that $a = (a \cup \{x\}) - \{x\} \in n$ by the first in our sequence of theorems here. This is the goal of the first part of the proof.

In the second part of the proof, we choose $a \in n$ arbitrarily and our goal is to show $a \in m$. The proof is precisely the same as the previous part with m and n interchanged.

So Axiom 4 of Peano arithmetic holds in our implementation.

A familiar construction of finite objects is the construction of *ordered pairs*.

***ordered pair:** We define $\langle x, y\rangle$ as $\{\{x\}, \{x, y\}\}$. Note that the pair is two types higher than its components x and y.

Theorem: For any x, y, z, w (all of the same type), $\langle x, y\rangle = \langle z, w\rangle$ iff $x = z$ and $y = w$.

Proof: This is left as an exercise.

***cartesian product:** For any sets A and B, we define $A \times B$, the *cartesian product* of A and B, as $\{\langle a, b\rangle \mid a \in A \wedge b \in B\}$. Notice that this is an example of generalized set builder notation, and could also be written as $\{c \mid (\exists ab.c = \langle a, b\rangle \wedge a \in A \wedge b \in B)\}$ (giving a promised example of the generalized set builder notation definition).

The definitions above are starred because we will in fact not use these common definitions. These definitions (due to Kuratowski) are usable in typed set theory and have in fact been used, but they have a practical disadvantage: the pair $\langle x, y\rangle$ is two types higher than its components x and y.

We will instead introduce a new primitive notion and axiom.

ordered pair: For any objects $x^{\mathbf{n}}$ and $y^{\mathbf{n}}$, we introduce primitive notation $\langle x^{\mathbf{n}}, y^{\mathbf{n}}\rangle^{\mathbf{n}}$ for the ordered pair of x and y and primitive notation $\pi_1(x^{\mathbf{n}})^{\mathbf{n}}$ and $\pi_2(x^{\mathbf{n}})^{\mathbf{n}}$ for the first and second projections of an object $x^{\mathbf{n}}$ considered as an ordered pair. As the notation suggests, the type of the pair is the same as the types of its components x and y (which we call its *projections*). In accordance with our usual practice, we will omit the type indices most of the time, allowing them to be deduced from the context.

Axiom of the Ordered Pair: For any x, $x = \langle \pi_1(x), \pi_2(x)\rangle$.

Corollary: For any x, y, z, w, $\langle x, y\rangle = \langle z, w\rangle \leftrightarrow x = z \wedge y = w$. The corollary is usually taken to be the defining property of the ordered pair; our axiom has the additional consequence that all objects are ordered pairs.

cartesian product: For any sets A and B, we define $A \times B$, the *cartesian product* of A and B, as $\{\langle a, b\rangle \mid a \in A \wedge b \in B\}$. Notice that this is an example of generalized set builder notation, and could also be written as $\{c \mid (\exists ab.c = \langle a, b\rangle \wedge a \in A \wedge b \in B)\}$ (giving a promised example of the generalized set builder notation definition).

We define A^2 as $A \times A$ and more generally define A^{n+1} as $A \times A^n$ (this definition of "cartesian powers" would not work if we were using the Kuratowski pair, for reasons of type). Notice that these exponents can be distinguished from type superscripts (when they are used) because we do not use boldface.

A crucial advantage of a type-level pair in practice is that it allows a nice definition of n-tuples for every n:

tuples: $\langle x_1, x_2, \ldots, x_n\rangle = \langle x_1, \langle x_2, \ldots, x_n\rangle\rangle$ for $n > 2$.

This would not type correctly if the Kuratowski pair were used, and an inelegant solution requiring the use of iterated singletons would be the best we could do along these lines.

We show that the Axiom of Infinity follows from the Axiom of Ordered Pairs (so we strictly speaking do not need the Axiom of Infinity if we assume the Axiom of Ordered Pairs).

Theorem: The Axiom of Ordered Pairs implies the Axiom of Infinity.

Proof: We argue that if $A \in n \in \mathbb{N}$ then $A \times \{0\} \in n$. $\emptyset$ is the only element of 0 and $\emptyset \times \{0\} = 0 \in \mathbb{N}$. Suppose that $A \times \{0\} \in n$ for all $A \in n$. Any element of $n+1$ is of the form $A \cup \{x\}$ where $A \in n$ and $x \notin A$. $(A \cup \{x\}) \times \{0\} = (A \times \{0\}) \cup \{\langle x, 0\rangle\} \in n+1$. The claim follows by induction. Now suppose $V \in N \in \mathbb{N}$. It follows that $V \times \{0\} \in N$. But certainly $V \times \{0\} \subseteq V$ so by a theorem about finite sets proved above, $V = V \times \{0\}$, which is absurd.

3.5.1 Digression: The Quine Ordered Pair

We develop a more complex definition of an ordered pair $\langle x, y\rangle$, due to Willard v. O. Quine, which is of the same type as its components x and y and satisfies the Axiom of Ordered Pairs above, but only works if strong extensionality is assumed.

The definition of the Quine pair is quite elaborate. The basic idea is that the Quine pair $\langle A, B\rangle$ is a kind of tagged union of A and B (it is only defined on sets of sets). Suppose that we can associate with each element a of A an object $\texttt{first}(a)$ from which a can be recovered, and with each element b of B an object $\texttt{second}(b)$ from which b can be recovered, and we can be sure that $\texttt{first}(a)$ and $\texttt{second}(b)$ will be distinct from each other for any $a \in A$ and $b \in B$. The idea is that $\langle A, B\rangle$ will be defined as

$$\{\texttt{first}(a) \mid a \in A\} \cup \{\texttt{second}(b) \mid b \in B\}.$$

For this to work we need the following things to be true for all objects x and y of the type to which elements of A and B belong:

1. For any x, y, $\texttt{first}(x) = \texttt{first}(y) \to x = y$
2. For any x, y, $\texttt{second}(x) = \texttt{second}(y) \to x = y$

3. For any x, y, $\texttt{first}(x) \neq \texttt{second}(y)$

If these conditions hold, then we can recover A and B from $\langle A, B\rangle$. An element x of $\langle A, B\rangle$ will be of the form $\texttt{first}(a)$ for some $a \in A$ or of the form $\texttt{second}(b)$ for some $b \in B$. It will be only one of these things, because no $\texttt{first}(x)$ is equal to any $\texttt{second}(y)$. Moreover, if $x = \texttt{first}(a)$, there is only one a for which this is true, and if $x = \texttt{second}(b)$ there is only one b for which this is true. So $A = \{a \mid \texttt{first}(a) \in \langle A, B\rangle\}$ and $B = \{b \mid \texttt{second}(b) \in \langle A, B\rangle\}$.

Thus if $\langle A, B\rangle = \langle C, D\rangle$ we have $A = \{a \mid \texttt{first}(a) \in \langle A, B\rangle\} = \{a \mid \texttt{first}(a) \in \langle C, D\rangle\} = C$ and similarly $B = D$.

The details of the definitions of the needed `first` and `second` operators follow. They will actually be called σ_1 and σ_2.

Definition: For each $n \in \mathbb{N}$ we define $\sigma(n)$ as $n + 1$ and for each $x \notin \mathbb{N}$ we define $\sigma(x)$ as x. Note that $\sigma(x)$ is of the same type as x.

Observation: For any x, y, if $\sigma(x) = \sigma(y)$ then $x = y$. If x and y are not natural numbers then this is obvious. If x is a natural number and y is not, then $\sigma(x)$ is a natural number and $\sigma(y)$ is not, so the hypothesis cannot be true. If x and y are natural numbers the statement to be proved is true by axiom 4.

Definition: We define $\sigma_1(x)$ as $\{\sigma(y) \mid y \in x\}$. We define $\sigma_2(x)$ as $\sigma_1(x) \cup \{0\}$. We define $\sigma_3(x)$ as $\{y \mid \sigma(y) \in x\}$. Note that all of these operations preserve type.

Observation: $\sigma_3(\sigma_1(x)) = x$, so if $\sigma_1(x) = \sigma_1(y)$ we have $x = \sigma_3(\sigma_1(x)) = \sigma_3(\sigma_1(y)) = y$; $\sigma_3(\sigma_2(x)) = x$, so similarly if $\sigma_2(x) = \sigma_2(y)$ we have $x = y$; $\sigma_1(x) \neq \sigma_2(y)$, because $0 \notin \sigma_1(x)$ and $0 \in \sigma_2(y)$. This shows that the σ_1 and σ_2 operations have the correct properties to play the roles of `first` and `second` in the abstract discussion above.

Definition: We define $\sigma_1\text{“}(x)$ as $\{\sigma_1(y) \mid y \in x\}$, $\sigma_2\text{“}(x)$ as $\{\sigma_2(y) \mid y \in x\}$ and $\sigma_3\text{“}(x)$ as $\{\sigma_3(y) \mid y \in x\}$

Definition: We define $\langle x, y\rangle$ as $\sigma_1\text{“}(x) \cup \sigma_2\text{“}(y)$. Note that the pair is of the same type as its components.

Theorem: For each set x there are unique sets $\pi_1(x)$ and $\pi_2(x)$ such that $\langle \pi_1(x), \pi_2(x) \rangle = x$. An immediate corollary is that for any x, y, z, w (all of the same type), $\langle x, y \rangle = \langle z, w \rangle$ iff $x = z$ and $y = w$.

Proof: $\pi_1(x) = \sigma_3\text{“}(\{y \in x \mid 0 \notin y\})$; $\pi_2(x) = \sigma_3\text{“}(\{y \in x \mid 0 \in y\})$

The Quine pair is defined only at type 4 and above; this is not a problem for us because we can do all our mathematical work in as high a type as we need to: notice that the natural numbers we have defined are present in each type above type 2; all mathematical constructions we present will be possible to carry out in any sufficiently high type.

In the theory with weak extensionality, the Quine pair is defined only on sets of sets (elements of $\mathcal{P}^{\mathbf{2}}(V)$) in types 4 and above, but it does satisfy the Axiom of Ordered Pairs on this restricted domain. We could in principle use the Quine pair instead of instroducing a primitive pair, if we were willing to restrict relations and functions to domains consisting of sets of sets. This isn't as bad as it seems because all objects of mathematical interest are actually sets of sets. We will not do this (our primitive pair acts on all objects), but we can use the Quine pair on sets of sets to justify our introduction of the primitive pair: if we cut down our universe to the sets of sets in types 4 and above, and use the relation $x \in' y$ defined as $x \in y \wedge y \in \mathcal{P}^{\mathbf{3}}(V)$ as our new membership relation (allowing only sets of sets of sets to be sets in the restricted world) it is straightforward to verify that our axioms will hold with the new membership relation and the Quine pair in the old world (with its associated projection functions) will still be a pair and projections in the new world satisfying the Axiom of Ordered Pairs. We can do even better. If we replace the natural numbers n in the definition of the Quine pair in the old world with $n \cap \mathcal{P}^{\mathbf{2}}(V)$, the pair in the new world will turn out to coincide with the new world's Quine pair on sets of sets (because the objects $n \cap \mathcal{P}^{\mathbf{2}}(V)$ are the natural numbers in the new world), and further all pairs of sets will be sets.

What we have just given is a sketch of what is called a *relative consistency proof*. Given a model of our type theory with the Axiom of Infinity, we show how to get a model of our type theory with the Axiom of Ordered Pairs (but not quite the same model).

Something important is going on here: we are forcibly reminded here that we are *implementing* already familiar mathematical concepts, not revealing what they "really are". Each implementation has advantages and

disadvantages. Here, the Kuratowski pair has the advantage of simplicity and independence of use of the Axiom of Infinity, while the Quine pair (or the primitive pair we have to introduce because we allow non-sets) has the technical advantage, which will be seen later to be overwhelming, that it is type level. Neither is the *true* ordered pair; the ordered pair notion prior to implementation is not any particular sort of object: its essence is perhaps expressed in the theorem that equal ordered pairs have equal components. The internal details of the implementation will not matter much in the sequel: what will do the mathematical work is the fact that the pair exactly determines its two components.

3.5.2 Exercises

1. Write a definition of the natural number 2 in the form $\{x \mid \phi[x]\}$ where ϕ is a formula containing only variables, logical symbols, equality and membership. Hint: the formula $\phi[x]$ needs to express the idea that x has exactly two elements in completely logical terms. How would you say that x has at least two elements? How would you say that x has at most two elements?

   A definition of 1 in this style is

   $$\{x \mid (\exists y.y \in x) \wedge (\forall uv.u \in x \wedge v \in x \rightarrow u = v)\}.$$

   Another definition of 1 is

   $$\{x \mid (\exists y.y \in x \wedge (\forall z.z \in x \rightarrow z = y))\}.$$

   Notice the different structure of the scopes of the quantifiers in the two definitions.

2. The usual definition of the ordered pair use in untyped set theory (due to Kuratowski) is

   $$\langle x, y\rangle =_{\mathtt{def}} \{\{x\}, \{x, y\}\}.$$

   We will not use this as our definition of ordered pair because it has the inconvenient feature that the pair is two types higher than its projections. What we *can* do (as an exercise in thinking about sets) is prove the following basic Theorem about this pair definition:

   $$\langle x, y\rangle = \langle z, w\rangle \rightarrow x = z \wedge y = w$$

   This is your exercise. There are various ways to approach it: one often finds it necessary to reason by cases. if you have seen a proof of this, don't go look it up: write your own.

3. Prove the theorem $(\forall xyz.\{x, z\} = \{y, z\} \rightarrow x = y)$ from the axioms of type theory, the definition of unordered pairs $\{u, v\}$, logic and the properties of equality. Remember that distinct letters do not necessarily represent distinct objects.

   This could be used to give a very efficient solution to the previous exercise.

4. Prove that the set $\mathbb{N}^{\mathbf{k+3}}$ (the set of natural numbers in type $\mathbf{k+3}$) is inductive. You don't need to specify types on every variable (or constant) every time it occurs, but you might want to state the type of each object mentioned in the proof the first time it appears.

   This proof is among other things an exercise in the careful reading of definitions.

5. Prove the following statement using the Peano axioms in the form stated in the current section: $(\forall n \in \mathbb{N}.n = 0 \vee (\exists m.m + 1 = n))$. You will need to use mathematical induction (in the set based form introduced above), but there is something very odd (indeed rather funny) about this inductive proof.

   Why is the object m unique in case it exists? (This is a throwaway corollary of the main theorem: it does not require an additional induction argument).

6. You are given that $n > 0$ is a natural number and a, b are not natural numbers.

   Compute the Quine pairs $\langle x, y\rangle$ and $\langle y, x\rangle$ where $x = \{\{\emptyset, 3\}, \{2\}, \{0, b\}\}$ and $y = \{\{1, 2\}, \{n, a\}\}$

   Given that $\langle u, v\rangle = \{\{0, 2, 4\}, \{a, b, 2\}, \{0\}, \{1\}, \{a, n\}\}$, what are the sets u and v?

7. Prove that the following are pair definitions (that is, show that they satisfy the defining theorem of ordered pairs).

   The Wiener pair: This is the first ordered pair definition in terms of set theory ever given.

   $$\langle x, y\rangle =_{\mathtt{def}} \{\{\{x\}\}, \{\{y\}, \emptyset\}\}.$$

   A pair that raises type by one: This is due to the author. Define $[x, a, b]$ as $\{\{x', a, b\} \mid x' \in x\}$. Define $\langle x, y\rangle$ as $[x, 0, 1] \cup [x, 2, 3] \cup [y, 4, 5] \cup [y, 6, 7]$, where 0,1,2,3,4,5,6,7 can be any eight distinct objects.

8. We define an *initial segment of the natural numbers* as a set S of natural numbers which has the property that for all natural numbers m, if $m + 1 \in S$ then $m \in S$.

Does an initial segment of the natural numbers need to contain all natural numbers? Explain why, or why not (with an example).

Prove that any nonempty initial segment of the natural numbers includes 0.

How do we prove *anything* about natural numbers?

9. Find sets A and B such that $A + 1 = B + 1$ but $A \neq B$. I found an example that isn't too hard to describe where $A + 1 = B + 1 = 3$ (or any large enough natural number; nothing special about 3). There are other classes of examples. This shows that Axiom 4 is true of natural numbers but not of sets in general.

We give some solutions.

2. We repeat the definition

$$\langle x, y\rangle =_{\mathtt{def}} \{\{x\}, \{x, y\}\}$$

of the Kuratowski pair. Our goal is to prove

$$(\forall xyzw.\, \langle x, y\rangle = \langle z, w\rangle \to x = z \wedge y = w).$$

We let x,y,z,w be arbitrarily chosen objects. Assume that $\langle x, y\rangle = \langle z, w\rangle$: our new goal is $x = z \wedge y = w$. Unpacking definitions tells us that we have assumed $\{\{x\}, \{x, y\}\} = \{\{z\}, \{z, w\}\}$.

We have two things to prove (since our goal is a conjunction). Note that these are not separate cases: the result proved as the first subgoal can (and will) be used in the proof of the second.

Goal 1: $x = z$

Proof of Goal 1: Because $\{\{x\}, \{x, y\}\} = \{\{z\}, \{z, w\}\}$, we have either $\{x\} = \{z\}$ or $\{x\} = \{z, w\}$. This allows us to set up a proof by cases.

Case 1a: We assume $\{x\} = \{z\}$. Certainly $x \in \{x\}$; thus by substitution $x \in \{z\}$, thus by definition of $\{z\}$ (and by comprehension) we have $x = z$.

Case 1b: We assume $\{x\} = \{z, w\}$. Certainly $z \in \{z, w\}$ (by definition of $\{z, w\}$ and comprehension). Thus $z \in \{x\}$, by substitution of equals for equals. Thus $z = x$, so $x = z$.

Conclusion: In both cases $x = z$ is proved, so Goal 1 is proved.

Goal 2: $y = w$

Proof of Goal 2: Note that we can use the result $x = z$ proved above in this subproof.

Because $\{\{x\}, \{x, y\}\} = \{\{z\}, \{z, w\}\}$ we have either $\{x\} = \{z, w\}$ or $\{x, y\} = \{z, w\}$. This allows us to set up an argument by cases.

Case 2a: Assume $\{x\} = \{z, w\}$. Since $z \in \{z, w\}$ and $w \in \{z, w\}$, we have $z \in \{x\}$ and $w \in \{x\}$ by substitution, whence we have $x = z = w$. This implies that $\{z\} = \{z, w\}$, so

$\{\{z\}, \{z, w\}\} = \{\{z\}\}$. Now we have $\{\{x\}, \{x, y\}\} = \{\{z\}\}$ by substitution into our original assumption, whence $\{x, y\} = \{z\}$, whence $x = y = z$ (the proofs of these last two statements are exactly parallel to things already proved), so $y = w$ as desired, since we also have $x = z = w$.

Case 2b: Assume $\{x, y\} = \{z, w\}$. Suppose $y \neq w$ for the sake of a contradiction. Since $y \in \{x, y\}$, we have $y \in \{z, w\}$, whence $y = z$ or $y = w$ Since $y \neq w$, we have $y = z$. Since $w \in \{x, y\}$ we have $w = x$ or $w = y$. Since $w \neq y$, we have $w = x$. Now we have $y = z = x = w$, so $y = w$, giving the desired contradiction, and completing the proof that $y = w$.

Conclusion: Since $y = w$ can be deduced in both cases, it can be deduced from our original assumption, completing the proof of Goal 2 and of the entire theorem.

5. Our goal is $(\forall n \in \mathbb{N}.n = 0 \vee (\exists m.m + 1 = n))$.

Define A as the set $\{n \in \mathbb{N} \mid n = 0 \vee (\exists m \in \mathbb{N}.m + 1 = n)\}$.

Our goal is to prove that A is inductive, from which it will follow that $\mathbb{N} \subseteq A$, from which the theorem follows.

Basis Step: $0 \in A$

Proof of Basis Step: $0 \in A \leftrightarrow (0 = 0 \vee (\exists m \in \mathbb{N}.m + 1 = 0))$, and $0 = 0$ is obviously true.

Induction Step: $(\forall k \in \mathbb{N}.k \in A \rightarrow k + 1 \in A)\}$.

Proof of Induction Step: Let k be an arbitrarily chosen natural number. Assume $k \in A$. Our goal is to prove $k + 1 \in A$, that is, $k+1 = 0 \vee (\exists m \in \mathbb{N}.m+1 = k+1)$. We prove this by observing that $k \in \mathbb{N}$ and $k+1 = k+1$, which witnesses $(\exists m \in \mathbb{N}.m+1 = k+1)$. Notice that the inductive hypothesis $k \in A$ was never used at all: there is no need to expand it.

7. We define an *initial segment of the natural numbers* as a set S of natural numbers which has the property that for all natural numbers m, if $m + 1 \in S$ then $m \in S$.

Does an initial segment of the natural numbers need to contain all natural numbers? Explain why, or why not (with an example).

Solution: No. The empty set is an initial segment, since the hypothesis $m + 1 \in S$ is false for every m if $S = \emptyset$, making $m + 1 \in S \rightarrow m \in S$ vacuously true. A nonempty initial segment not equal to $\mathbb{N}$ is for example $\{0, 1\}$: the implication can be checked for $m = 0$ and is vacuously true for all other values of m.

Prove that any nonempty initial segment of the natural numbers includes 0.

Solution: Let S be a nonempty initial segment of the natural numbers. Our goal is to show $0 \in S$. Since S is nonempty, we can find $m \in S$. If we could show $(\forall n \in \mathbb{N}.n \in S \rightarrow 0 \in S)$, we would have $m \in S \rightarrow 0 \in S$ and $0 \in S$ by modus ponens.

We prove the lemma $(\forall n \in \mathbb{N}.n \in S \rightarrow 0 \in S)$ by mathematical induction. Let $A = \{n \in \mathbb{N} \mid n \in S \rightarrow 0 \in S\}$. We show that A is inductive.

Basis Step: $0 \in S \rightarrow 0 \in S$ is the goal. This is obvious.

Induction Step: Let k be an arbitrarily chosen natural number. Suppose $k \in A$. Our goal is $k + 1 \in A$. $k \in A$ means $k \in A \rightarrow 0 \in S$. We have $k+1 \in S \rightarrow k \in S$ because S is an initial segment. From these two implications $k + 1 \in S \rightarrow 0 \in S$ follows, completing the proof of the induction step and the lemma.

3.6 Relations and Functions

If A and B are sets, we define a *relation from A to B* as a subset of $A \times B$. A *relation* in general is simply a set of ordered pairs.

If R is a relation from A to B, we define $x\,R\,y$ as $\langle x, y\rangle \in R$. This notation should be viewed with care. Note here that x and y must be of the same type, while R is one type higher than x or y (that would be three types higher if we used the Kuratowski pair). In the superficially similar notation $x \in y$, y is one type higher than x and $\in$ does not denote a set at all: do not confuse logical relations with set relations. In some cases they can be conflated: the notation $x \subseteq y$ can be used to motivate a definition of $\subseteq$ as a set relation ($[\subseteq] = \{\langle x, y\rangle \mid x \subseteq y\}$), though we do not originally understand $x \subseteq y$ as saying anything about a set of ordered pairs.

If R is a relation, we define $\mathtt{dom}(R)$, the *domain of R*, as $\{x \mid (\exists y.x\,R\,y)\}$. We define R^{-1}, the *inverse of R*, as $\{\langle x, y\rangle \mid y\,R\,x\}$. We define $\mathtt{rng}(R)$, the *range of R*, as $\mathtt{dom}(R^{-1})$. We define $\mathtt{fld}(R)$, the *field of R*, as the union of $\mathtt{dom}(R)$ and $\mathtt{rng}(R)$. If R is a relation from A to B and S is a relation from B to C, we define $R|S$, the *relative product of R and S* as $\{\langle x, z\rangle \mid (\exists y.x\,R\,y \wedge y\,S\,z)\}$.

The symbol $[=]$ is used to denote the equality relation $\{\langle x, x\rangle \mid x \in V\}$. Similarly $[\subseteq]$ can be used as a name for the subset relation (as we did above), and so forth: the brackets convert a grammatical "transitive verb" to a noun.

We define special characteristics of relations.

reflexive: R is *reflexive* iff $x\,R\,x$ for all $x \in \mathtt{fld}(R)$.

symmetric: R is *symmetric* iff for all x and y, $x\,R\,y \leftrightarrow y\,R\,x$.

antisymmetric: R is *antisymmetric* iff for all x, y if $x\,R\,y$ and $y\,R\,x$ then $x = y$.

asymmetric: R is *asymmetric* iff for all x, y if $x\,R\,y$ then $\neg y\,R\,x$. Note that this immediately implies $\neg x\,R\,x$.

transitive: R is *transitive* iff for all x, y, z if $x\,R\,y$ and $y\,R\,z$ then $x\,R\,z$.

equivalence relation: A relation is an *equivalence relation* iff it is reflexive, symmetric, and transitive.

partial order: A relation is a *partial order* iff it is reflexive, antisymmetric, and transitive.

strict partial order: A relation is a *strict partial order* iff it is asymmetric and transitive. Given a partial order R, $R - [=]$ will be a strict partial order. From a strict partial order $R - [=]$, the partial order R can be recovered if it has no "isolated points" (elements of its field related only to themselves).

linear order: A partial order R is a *linear order* iff for any $x, y \in \mathtt{fld}(R)$, either $x\,R\,y$ or $y\,R\,x$. Note that a linear order is precisely determined by the corresponding strict partial order if its domain has two or more elements.

strict linear order: A strict partial order R is a *strict linear order* iff for any $x, y \in \mathtt{fld}(R)$, one has $x\,R\,y$, $y\,R\,x$ or $x = y$. If R is a linear order, $R - [=]$ is a strict linear order.

image: For any set $A \subseteq \mathtt{fld}(R)$, $R\text{“}A = \{b \mid (\exists a \in A.a\,R\,b)\}$.

extensional: A relation R is said to be *extensional* iff for any $x, y \in \mathtt{fld}(R)$, $R^{-1}\text{“}(\{x\}) = R^{-1}\text{“}(\{y\}) \rightarrow x = y$: elements of the field of R with the same preimage under R are equal. An extensional relation supports a representation of some of the subsets of its field by the elements of its field.

well-founded: A relation R is *well-founded* iff for each nonempty subset A of $\mathtt{fld}(R)$ there is $a \in A$ such that for no $b \in A$ do we have $b\,R\,a$ (we call this a minimal element of A with respect to R, though note that R is not necessarily an order relation).

well-ordering: A linear order R is a *well-ordering* iff the corresponding strict partial order $R - [=]$ is well-founded.

strict well-ordering: A strict linear order R is a *strict well-ordering* iff it is well-founded.

end extension: A relation S *end extends* a relation R iff $R \subseteq S$ and for any $x \in \mathtt{fld}(R)$, $R^{-1}\text{“}\{x\} = S^{-1}\text{“}\{x\}$. (This is a nonstandard adaptation of a piece of terminology from model theory).

function: f is a *function from A to B* (written $f : A \rightarrow B$) iff f is a relation from A to B and for all x, y, z, if $x\,f\,y$ and $x\,f\,z$ then $y = z$. For each $x \in \mathtt{dom}(f)$, we define $f(x)$ as the unique y such that $x\,f\,y$ (this exists

because x is in the domain and is unique because f is a function). The notation $f[A]$ is common for the image $f\text{“}A$.

warning about function notation: Notations like $\mathcal{P}(x)$ for the power set of x should not be misconstrued as examples of the function value notation $f(x)$. There is no function $\mathcal{P}$ because $\mathcal{P}(x)$ is one type higher than x. We have considered using the notation $F\text{`}x$ (this was Russell's original notation for function values) for defined operators in general and restricting the notation $f(x)$ to the case where $f$ is actually a set function. If we did this we would exclude (for example) the notation $\mathcal{P}(x)$ in favor of $\mathcal{P}\text{`}x$ (or $\mathcal{P}\text{`}(t)$ for complex terms $t$ that require parentheses). If we used the Russell notation in this way we would also write $\bigcup\text{`}x, \bigcap\text{`}x$ because these operations also shift type. We would then prefer the use of $f[A]$ to the use of $f\text{“}A$ for images under functions. But we have not adopted such a convention here.

injection: A function f is an *injection* (or *one-to-one*) iff f^{-1} is a function.

surjection: A function f is a *surjection from A to B* or a *function from A onto B* iff it is a function from A to B and $f\text{“}A = B$.

bijection: A function f is a *bijection from A to B* iff it is an injection and also a surjection from A to B.

composition and restriction: If f is a function and A is a set (usually a subset of $\mathtt{dom}(f)$), define $f\restriction A$ as $f \cap (A \times V)$ (the *restriction of f to the set A*). If f and g are functions and $\mathtt{rng}(g) \subseteq \mathtt{dom}(f)$, define $f \circ g$ as $g|f$. This is called the *composition* of f and g. Because the order of composition is unnatural, we will often write compositions as relative products.

identity function: Note that $[=]$ is a function. We call it the *identity function*, and we call $[=]\restriction A$ the *identity function on A*, where A is any set.

abstraction: If $T[x]$ is a term (usually involving x) define $(x : A \mapsto T[x])$ or $(\lambda x : A.T[x])$ as $\{\langle x, T[x]\rangle \mid x \in A\}$. The explicit mention of the set A may be omitted when it is V or when it is understood from the form of the term $T[x]$.

3.6.1 Exercises

1. I give definitions of injective and surjective function from A to B (not identical to those in the book, though you are welcome to verify that they are equivalent).

   A function f is an injective function from A to B iff it is a function from A to B and for all $x, y \in A$, $f(x) = f(y) \rightarrow x = y$.

   A function f is a surjective function from A to B iff it is a function from A to B and for all $y \in B$, there exists $x \in A$ such that $f(x) = y$.

   Prove that if f is an injective function from A to B and g is an injective function from B to C, then $g \circ f$ is an injective function from A to C. ($g \circ f$ may be supposed defined by the equation $(g \circ f)(x) = g(f(x))$).

   Prove that if f is an surjective function from A to B and g is a surjective function from B to C, then $g \circ f$ is an surjective function from A to C.

   Use the definitions given in the problem and proof strategy as described in section 2.

   Comment: of course this shows compositions of bijections are bijections, which will be useful.

3.7 Defining Functions by Recursion; First-Order Peano Arithmetic

Recursion is a special technique for defining functions with domain $\mathbb{N}$.

Informally, a recursive definition might look like this (this is not a completely general example): $f(0) = 0$; for each natural number n, $f(n+1) = (f(n) + 1) + 1$. This seems somehow suspect because this definition of f appears to mention f itself in an essential way.

We show that this kind of definition is legitimate. We begin by exhibiting the technique of *iterative* definition of which the example just given is a special case.

Iteration Theorem: For any set a and function f (of appropriate types) there is a unique function $g : \mathbb{N} \rightarrow V$ such that $g(0) = a$ and $g(n+1) = f(g(n))$ for each $n \in \mathbb{N}$.

Definition: Where a, f, g are as in the statement of the Theorem, we define $f^n(a)$ as $g(n)$.

Proof of Iteration Theorem: We begin with a nonce

> **Definition:** A set I is said to be (f, a)-*inductive* iff $\langle 0, a\rangle \in I$ and $(\forall nx.\, \langle n, x\rangle \in I \rightarrow \langle n+1, f(x)\rangle \in I)$.

Let g be the intersection of all (f, a)-inductive sets. We claim that g is the desired function. Note that we do not even know that g is a function at this point!

We claim that g is a subset of $\mathbb{N} \times V$. Note that $\langle 0, a\rangle \in \mathbb{N} \times V$ and for any $\langle n, x\rangle \in \mathbb{N} \times V$ we also have $\langle n+1, f(x)\rangle \in \mathbb{N} \times V$, so $\mathbb{N} \times V$ is (f, a)-inductive, whence $g \subseteq \mathbb{N} \times V$.

So we now know that every element of g is an ordered pair whose first component is a natural number, which is necessary but not sufficient for g to be a function with domain the set of natural numbers.

We claim that for each natural number n there is exactly one object x such that $\langle n, x\rangle$ is an element of g. Define A as the set of all natural numbers n such that there is exactly one object x such that $\langle n, x\rangle$ is an element of g: we prove our claim by showing that A is inductive.

We first need to show that $0 \in A$. We know that $\langle 0, a\rangle \in g$, so there is at least one x such that $\langle 0, x\rangle \in g$. Now consider $g' = g - \{\langle 0, x\rangle \mid x \neq a\}$. We claim that g' is (f, a)-inductive. $\langle 0, a\rangle \in g'$ is obvious. Suppose $\langle n, x\rangle \in g'$. It follows that $\langle n+1, f(x)\rangle \in g$, and in fact that $\langle n+1, f(x)\rangle \in g'$, because $\langle n+1, f(x)\rangle \notin \{\langle 0, x\rangle \mid x \neq a\}$. Since g' is (f, a)-inductive, $g \subseteq g'$. But $g' \subseteq g$ as well, so $g = g'$, and a is the only object such that $\langle 0, a\rangle \in g' = g$, which is what we needed to show.

Now we need to show that for any $k \in A$ we also have $k+1 \in A$. Assume $k \in A$, whence there is exactly one u such that $\langle k, u\rangle \in g$. We need to show that there is exactly one v such that $\langle k+1, v\rangle \in g$. Since $\langle k, u\rangle \in g$, it follows that $\langle k+1, f(u)\rangle \in g$, so there is at least one such v. Now define g' as $g - \{\langle k+1, w\rangle \mid w \neq f(u)\}$. We claim that g' is (f, a)-inductive. Clearly $\langle 0, a\rangle \in g'$. Suppose $\langle n, x\rangle \in g'$; our aim is to show $\langle n+1, f(x)\rangle \in g'$. Suppose otherwise for the sake of a contradiction. Clearly $\langle n+1, f(x)\rangle \in g$: it is thus necessary that $\langle n+1, f(x)\rangle \in \{\langle k+1, w\rangle \mid w \neq f(u)\}$, which implies $f(x) \neq f(u)$ and also that $n+1 = k+1$. From this it follows that $n = k$, and thus, since $\langle n, x\rangle = \langle k, x\rangle \in g$, that $x = u$, whence $f(x) \neq f(u)$ is impossible,

which is the desired contradiction. We then have $g = g'$, whence $f(u)$ is the only object x such that $\langle k+1, x\rangle \in g' = g$, whence $k+1 \in A$.

This completes the proof that g is a function from $\mathbb{N}$ to V. Since $\langle 0, a\rangle \in g$, we have $g(0) = a$. Since $\langle n, g(n)\rangle \in g$, we have $\langle n+1, f(g(n))\rangle \in g$, whence $g(n+1) = f(g(n))$.

Now we need to show that g is the unique function with these properties. Suppose $g' : \mathbb{N} \to V$, $g'(0) = a$ and $g'(n+1) = f(g'(n))$. $\langle 0, a\rangle \in g'$ is immediate. If $\langle n, x\rangle \in g'$, then $x = g'(n)$, and $\langle n+1, g'(n+1)\rangle = \langle n+1, f(g'(n))\rangle = \langle n+1, f(x)\rangle \in g'$, so g' is (f, a)-inductive, whence $g \subseteq g'$. g' contains exactly one element with first projection n for each natural number n, which must be the one element with first projection n belonging to g, so g and g' are the same set.

This completes the proof of the Iteration Theorem.

Recursion Theorem: For any set a and function $g : (\mathbb{N} \times V) \to V$, there is a function $h : \mathbb{N} \to V$ such that $h(0) = a$ and $h(n+1) = g(n, h(n))$ for each $n \in \mathbb{N}$.

Proof of Recursion Theorem: Let $G(\langle n, x\rangle)$ be defined as $\langle n+1, g(n)\rangle$. Then $h(n) = \pi_2(G^n(\langle 0, a\rangle))$.

There is an alternative way to define $f^n(a)$.

Definition: A set S of natural numbers is an *initial segment of the natural numbers* iff for all $n \in \mathbb{N}$, $n+1 \in S \to n \in S$.

Theorem: Any nonempty initial segment of the natural numbers contains 0.

Theorem: $y = f^n(a)$ iff there is a function g such that the domain of g is an initial segment S of the natural numbers including n as an element, $g(0) = a$, for all m such that $m+1 \in S$ we have $g(m+1) = f(g(m))$, and $y = g(n)$. This formulation is advantageous because it only appeals to the existence of finite sets.

As examples we can present definitions of addition and multiplication.

Give the nonce name σ to the successor function. We can define $m+n$ (for any natural numbers m, n) as $\sigma^n(m)$ (adding n is iterating successor n

times). We can define $m \cdot n$ for any natural numbers m and n as $(\sigma^m)^n(0)$: to add $m \cdot n$ is to add m n times.

The recursive (really as we see above "iterative") definitions of addition and multiplication are incorporated into modern formulations of "Peano's axioms", which make no essential reference to sets. The theory with these axioms is formally called *first-order Peano arithmetic*.

When we reason in first-order Peano arithmetic, we are not reasoning in our type theory. But, since we have shown that there is an intepretation of the axioms of first-order Peano arithmetic in our type theory, any theorems we prove in first-order Peano arithmetic will be true in that interpretation. We will see below that there is a different interpretation of Peano arithmetic commonly used in untyped set theory (the von Neumann definition of the natural numbers, already mentioned above), and anything we prove in arithmetic will also be true in that interpretation (and in any other we come up with).

The convention when reasoning in first-order Peano arithmetic is to assume that all quantifiers are restricted to the natural numbers (we are not talking about anything else, and notably we are not talking about sets of natural numbers as we do in the original (second-order) version of the theory). Note this particularly in axiom 9.

1. 0 is a natural number.
2. For each natural number n, $\sigma(n)$ is a natural number.
3. For all natural numbers n, $\sigma(n) \neq 0$
4. For all natural numbers m, n, $\sigma(m) = \sigma(n) \to m = n$.
5. For all natural numbers m, n, $m + 0 = m$
6. For all natural numbers m, n, $m + \sigma(n) = \sigma(m + n)$
7. For all natural numbers m, n, $m \cdot 0 = 0$
8. For all natural numbers m, n, $m \cdot \sigma(n) = m \cdot n + m$
9. For each formula $\phi[n]$, we adopt as an axiom $\phi[0] \wedge (\forall k.\phi[k] \to \phi[\sigma(k)]) \to (\forall n.\phi[n])$. This is the principle of mathematical induction. Note that this is not really a single axiom 9, but a suite of axioms 9_ϕ. Such a

suite is called an *axiom scheme*. A scheme is needed because we do not refer to sets here.

Since addition is also a primitive operation here, we use a primitive notation for successor at first rather than the more natural addition of 1. Notice the reformulation of mathematical induction in terms of formulas rather than sets. This formulation of mathematical induction is not a statement with a quantifier over formulas (we cannot really do that for reasons which we may discuss *much* later on) but an infinite collection of different axioms, one for each formula ϕ. You should notice that the axioms for addition and multiplication capture the iterative definitions of addition and multiplication given above.

We give some sample proofs in Peano arithmetic.

Definition: $1 = \sigma(0)$. Note that it is immediate from the axioms for addition that $n+1 = n+\sigma(0) = \sigma(n+0) = \sigma(n)$. We feel free to use these notations interchangeably.

Proof Strategy: We give the first order version of mathematical induction as a proof strategy.

To deduce a goal $(\forall n.\phi[n])$, deduce the following two goals:

Basis step: Deduce $\phi[0]$.

Induction step: Deduce $(\forall k.\phi[k] \rightarrow \phi[k+1])$. Application of prior proof strategy expands this: let k be an arbitrarily chosen natural number (which might be 0!): assume $\phi[k]$ (this is called the *inductive hypothesis*, and it is useful to emphasize where in an induction proof the inductive hypothesis is used), and deduce the new goal $\phi[k+1]$.

Theorem: For each natural number $n \neq 0$, there is a unique natural number m such that $m + 1 = n$.

Proof: We prove by mathematical induction the assertion "For each natural number n, if $n \neq 0$, then there is a natural number m such that $m+1 = n$".

For $n = 0$ this is trivially true (basis step).

Suppose it is true for $n = k$; then our goal is to prove that it is true for $n = k + 1$ (induction step).

Either $k = 0$ or there is an m such that $m + 1 = k$, by inductive hypothesis. In either case, there is an m' such that $m' + 1 = k + 1$, namely k itself.

So the assertion is true for all n by mathematical induction. What is strange here is that the inductive hypothesis is not used in this proof!

The observant reader will notice that we have not yet proved the theorem. We have shown that for each nonzero natural number n there is an m such that $m + 1 = n$, but we have not shown that this m is unique yet. Suppose that $m + 1 = n$ and also $m' + 1 = n$: it follows directly from an axiom that $m = m'$. So we have shown that there can only be one such m for each n and the proof is complete.

Theorem: For each natural number n, $0 + n = n + 0$.

Proof: We prove this by mathematical induction.

$0 + 0 = 0 + 0$ completes the proof of the basis step.

Now for the induction step. We assume that $0 + k = k + 0$ and our goal is to show that $0 + \sigma(k) = \sigma(k) + 0$. $0 + \sigma(k) = \sigma(0 + k)$ by axioms, and $\sigma(0 + k) = \sigma(k + 0)$ (by inductive hypothesis) $= \sigma(k) = \sigma(k) + 0$. This completes the proof of the induction step and of the theorem.

Theorem: For any natural numbers m, n, $(m + 1) + n = (m + n) + 1$.

We fix m and prove this by induction on n.

The basis step is established by $(m + 1) + 0 = m + 1 = (m + 0) + 1$.

The hypothesis of the induction step is $(m + 1) + k = (m + k) + 1$; the goal is to show $(m+1)+(k+1) = (m+(k+1))+1$. $(m+1)+(k+1) = ((m + 1) + k) + 1$ by axiom, which is equal to $((m + k) + 1) + 1$ by inductive hypothesis, which is in turn equal to $(m + (k + 1)) + 1$ by axiom, completing the proof.

Theorem: For any natural numbers m, n, $m + n = n + m$.

Proof: We prove this by (you guessed it!) mathematical induction.

The statement we actually prove by mathematical induction is "for any natural number n, for any natural number m, $m + n = n + m$."

The basis step is "For any natural number m, $m + 0 = 0 + m$". We just proved that!

The induction hypothesis is "For any natural number m, $m+k = k+m$" (for some fixed natural number k) and the induction goal is "For any natural number m, $m + (k + 1) = (k + 1) + m$". Now $m + (k + 1) = (m+k)+1$ by axiom, which is in turn equal to $(k+m)+1$ by inductive hypothesis, which is equal to $(k + 1) + m$ by the previous theorem, proving the induction goal and completing the proof of the theorem.

Much more natural definitions of the arithmetic operations which use the intuitive idea that the numbers are sizes of sets are given below, and in terms of these definitions much more natural proofs of properties such as the ones just proved can be given. Proofs in Peano arithmetic are nonetheless a useful exercise: they apply to quite different implementations of the natural numbers (another implementation will be given later): for any implementation, if the Peano axioms hold, then all the theorems following from the Peano axioms also hold.

Apparently stronger forms of both induction and recursion are available, but turn out to be equivalent to the basic forms already given. A presentation of these requires some prior discussion of the familiar order on the natural numbers.

Definition: For natural numbers m, n, we say $m \leq n$ (*m is less than or equal to n*) just in case $(\exists k.m + k = n)$. We define $m < n$ (*m is less than n*) as $m \leq n \wedge m \neq n$. We define $m \geq n$ (*m is greater than or equal to n*) as $n \leq m$, and similarly define $m > n$ (*m is greater than n*) as $n < m$.

Note that we assume here that such things as the associative and commutative laws of addition have already been proved.

Theorem: For all natural numbers m, n, k, if $m + k = n + k$ then $m = n$.

Proof: Fix m and n and prove by induction on k. This is obvious for $k = 0$. If it is true for k and $m+(k+1) = n+(k+1)$, then $(m+k)+1 = (n+k)+1$ by addition axiom, $m+k = n+k$ by axiom 4, and $m = n$ by inductive hypothesis.

Theorem: The relation $\leq$ on natural numbers just defined is a linear order.

Proof: $n \leq n = n + 0$ is immediate. If $m \leq n$ and $n \leq m$ then we have $n = m+k$ and $m = n+l$ for some k and l, whence $n = n+0 = n+(k+l)$, so $k + l = 0$, whence it is easy to show that $k = l = 0$, so $m = n$. If $m \leq n$ and $n \leq p$, then for some k, l, $m + k = n$ and $n + l = p$, so $(m + k) + l = m + (k + l) = p$.

Theorem: $m \leq n \leftrightarrow m + k \leq n + k$.

$m + p = n \leftrightarrow (m + k) + p = n + k$

Corollary: $m < n \leftrightarrow m + k < n + k$

Theorem: For all $n \in \mathbb{N}$, for all $k \in \mathbb{N}$, $k \leq n \leftrightarrow k < n + 1$.

Proof: Prove this by induction on n. The basis step requires us to show that $m \leq 0 \leftrightarrow m < 1$ for all m. If $m \leq 0$, then since $0 \leq 1$ and $1 \not\leq 0$, $m < 1$ is obvious. If $m \neq 0$ then $m = n + 1$ for some n, so $1 \leq m$, thus $m < 1 \rightarrow m = 0$ (by contrapositive). Now if $m \leq k \leftrightarrow m < k + 1$, for all m, we immediately have $(m+1) \leq (k+1) \leftrightarrow (m+1) < (k+1)+1$ We certainly also have $0 \leq k + 1 \leftrightarrow 0 < (k + 1) + 1$, and since every number is either 0 or a successor we have shown for all m that $m \leq k + 1 \leftrightarrow m < (k + 1) + 1$

Theorem (Strong Induction, set form): For any set A of natural numbers, if $(\forall a \in \mathbb{N}.(\forall x < a.x \in A) \rightarrow a \in A)$, then $A = \mathbb{N}$.

Proof: Suppose that A is a set of natural numbers and $(foralla \in \mathbb{N}.(\forall x < a.x \in A) \rightarrow a \in A)$. We define the set B as $\{b \in \mathbb{N}.(\forall x \leq B.x \in A)\}$. We show that B is inductive. Since $B \subseteq A$ is obvious, $B = \mathbb{N} \rightarrow A = \mathbb{N}$.

Since $(\forall x < 0.x \in A)$ is vacuously true, $0 \in A$. For any $b \leq 0$, $b = 0 \in A$, so $0 \in B$.

Now suppose that $k \in B$. Our goal is to show that $k + 1 \in B$. Since $k \in B$, we have $p \in A$ for all $p \leq k$, and so for all $p < k + 1$. It then follows that $k+1 \in A$, and since we have $p \in A$ for all $p < k+1$ as well, we also have $k + 1 \in B$. This completes the proof that B is inductive, which we have already seen is sufficient for the proof of the theorem.

Theorem (Strong Induction, property form): For any formula ϕ, $(\forall a \in \mathbb{N}.(\forall x < a.\phi[x]) \rightarrow \phi[a]) \rightarrow (\forall n \in \mathbb{N}.\phi[n])$.

Proof: This is proved in the same way as the previous theorem.

There is a form of recursion which is to standard recursion (or iteration) roughly as strong induction is to standard induction.

Theorem (Course-of-Values Recursion): Let A be a set. Let $\mathcal{F}$ be the set of all functions with domain a proper initial segment

$$\{m \in \mathbb{N} \mid m < n\}$$

of the natural numbers and range a subset of A (notice that the function with domain $\emptyset$ is one of these: set $n = 0$). Let G be any function from $\mathcal{F}$ to $\iota``A$. Then there is a uniquely determined function $f : \mathbb{N} \to A$ such that $f(n) \in G(f\lceil\{m \in \mathbb{N} \mid m < n\})$ for each $n \in \mathbb{N}$.

Proof: We define a function H from $\mathcal{F}$ to $\mathcal{F}$ as follows. If $g \in \mathcal{F}$ has domain $\{m \in \mathbb{N} \mid m < n\}$, define $H(g)$ as $g \cup (\{n\} \times G(g))$ (recall that $G(n)$ is the singleton set containing the intended value at n of the function being constructed). Now apply the iteration theorem: define $f(n)$ as $H^n(\emptyset)(n)$. It is straightforward to verify that this function has the desired property.

Example: An example of a function defined in this way, in which the value of f at any natural number depends on its values at *all* smaller natural numbers, would be $f(n) = 1 + \Sigma_{i<n} f(i)$

It is a usual exercise in a book of this kind to prove theorems of Peano arithmetic up to the point where it is obvious that the basic computational axioms of arithmetic and algebra can be founded on this basis (and we may do all of this in these notes or in exercises). It is less obvious that all usual notions of arithmetic and algebra can actually be defined in terms of the quite restricted vocabulary of Peano arithmetic and logic: this is very often asserted but seldom actually demonstrated. We supply an outline of how this can be established.

We give basic definitions without (or with only an indication of) supporting proofs to indicate that the expressive power of Peano arithmetic without set language is enough to talk about finite sets of natural numbers and to define recursive functions. This is a serious question because the definition of recursive functions above relies strongly on the use of sets. Notice that we

use the alternative formulation of the definition of $f^n(a)$ in this development, because we only code finite sets of natural numbers as natural numbers here, and the alternative formulation has the advantage that it only talks about finite sets.

Definition: For natural numbers m, n we say $m|n$ (n *is divisible by* m or m *is a factor of* n) iff there is a natural number x such that $m \cdot x = n$.

Definition: A natural number p is a *prime* iff it has exactly two factors. (One of these factors must be 1 and the other $p \neq 1$ itself).

Definition: Let p be a prime. A natural number q is a *power of* p iff p is a factor of every factor of q except 1.

Definition: Let p be a prime and n a natural number. A nonzero natural number m *occurs in the base* p *expansion of* n just in case n can be expressed in the form $a \cdot q + m \cdot r + s$, where $q > r > s$ and q, r are powers of p.

The underlying idea is that we now have the ability to code finite sets of natural numbers as natural numbers (and so in fact sets of sets, sets of sets of sets, and so forth).

Definition: Define $x \in_p y$ as "$x + 1$ occurs in the base p expansion of y". For any prime p and naturals $x_1, \ldots, x_n$ all less than $p - 1$ define $\{x_1, \ldots, x_n\}_p$ as the smallest natural number y such that $(\forall z. z \in_p y \leftrightarrow z = x_1 \vee \ldots \vee z = x_n)$. [there is something to prove here, namely that there is such a y].

Definition: Define $\langle x, y \rangle_{p,q}$ as $\{\{x\}_p, \{x, y\}_p\}_q$.

Definition: For any function f, we say that f is *definable in Peano arithmetic* iff there is a formula $\phi[x, y]$ in the language of arithmetic such that $\phi[x, y] \leftrightarrow y = f(x)$.

Theorem: For any function f definable in Peano arithmetic, $y = f^n(x)$ iff there are primes $p < q < r$ such that there is a natural number g such that $(\forall m \leq n.(\exists! y. \langle m, y \rangle_{p,q} \in_r g))$ and $\langle 0, x \rangle \in_r g$ and $(\forall m < n.(\forall y. \langle m, y \rangle_{p,q} \in_r g \rightarrow \langle m + 1, f(y) \rangle_{p,q} \in_r g))$. Note that this is expressible in the language of Peano arithmetic, so all functions definable by iteration of definable functions are definable (and functions

definable by recursion from definable functions are also definable since we can represent pairs of natural numbers as natural numbers and define the projection functions of these pairs).

Definition: Define $d(x)$ as $2 \cdot x$. Define 2^n as $d^n(1)$. Define $x \in_{\mathbb{N}} a$ as

$$(\exists y > x.(\exists z < 2^x.(\exists u.a = u \cdot 2^y + 2^x + z))).$$

This expresses that the nth digit in the binary expansion of a is 1, and this supports a nice coding of finite sets of natural numbers as natural numbers, which we will have occasion to use later.

3.7.1 Exercises

1. If I define a function I_n such that $I_n(f) = f^n$ (so for example $I_3(f)(x) = f^3(x) = f(f(f(x)))$, I invite you to consider the functions $(I_n)^m$. For example, compute $(I_2)^3(f)(x)$. Compute $(I_3)^2(f)(x)$. There is an equation $(I_m)^n = I_{F(m,n)}$, where F is a quite familiar operation on natural numbers, which you can write and might derive if you do enough experiments. There is a serious formal problem with this equation, though, in our type theory. What is the function $F(m, n)$? What is the formal problem?

2. Prove the theorem

   $$(\forall mn.(m + 1) + n = (m + n) + 1)$$

   of Peano arithmetic.

   Indicate each application of an axiom and of an inductive hypothesis. Do not apply theorems you have not proved yourself on your paper. You may identify $\sigma(x)$ and $x + 1$ without comment for any natural number x.

3. Prove as many of the following as you can in first-order Peano arithmetic, not necessarily in the given order. Your proofs should not mention sets or the type theory definitions of the natural numbers (this is all just arithmetic from the Peano axioms).

   Use proof strategy. You can be a little more freeform than heretofore, but take pains to make it clear what you are doing. You may use

theorems already proved in the notes or already proved by you. You may *not* use anything else you think you know about arithmetic.

Do prove at least two of them.

The associative law of addition.

The commutative law of multiplication.

The associative law of multiplication.

The distributive law of multiplication over addition.

3.8 Equivalence Relations, Partitions, and Representatives: the Axiom of Choice

Definition: Sets A and B are said to be *disjoint* just in case $A \cap B = \emptyset$.

Definition: A collection P of sets is said to be *pairwise disjoint* just in case

$$(\forall A \in P.(\forall B \in P.A = B \vee A \cap B = \emptyset)).$$

Definition: A collection P of sets is a *partition of* A iff $\emptyset \notin P$, $\bigcup P = A$, and P is pairwise disjoint. A partition of A is a collection of nonempty sets which do not overlap and which cover all of A. We say that a collection P *is a partition* iff it is a partition of $\bigcup P$.

Definition: If R is an equivalence relation and $x \in \texttt{fld}(R)$ we define $[x]_R$, the *equivalence class of* x *under* R, as $R\text{“}(\{x\}) = \{y \mid x\,R\,y\}$.

Theorem: If R is an equivalence relation, $P_R = \{[x]_R \mid x \in \texttt{fld}(R)\}$ is a partition of $\texttt{fld}(R)$.

Proof: Let R be an arbitrarily chosen equivalence relation. Define $P_R = \{[x]_R \mid x \in \texttt{fld}(R)\}$.

Our goal is to prove that P_R is a partition of $\texttt{fld}(R)$. Using the definition of partition, this reduces to three subgoals.

Goal 1: $\emptyset \notin P_R$**.** Suppose for the sake of a contradiction that $\emptyset \in P_R$. By the definition of P_R as a complex set abstract, this is equivalent to the assertion that $\emptyset = [x]_R$ for some $x \in \texttt{fld}(R)$. Choose such an x. $x\,R\,x$ holds because R is reflexive, whence $x \in [x]_R$ by the

definition of equivalence class, whence $x \in \emptyset$, which yields the desired contradiction. This completes the proof of Goal 1.

Goal 2: $\bigcup P_R = \texttt{fld}(R)$**.** Use the proof strategy for showing the equality of two sets.

2a: Let x be an arbitrarily chosen element of $\bigcup P_R$: our new goal is to show $x \in \texttt{fld}(R)$. Since $x \in \bigcup P_R$, we can choose a set A such that $x \in A$ and $A \in P_R$. Since $A \in P_R$, we can choose y such that $A = [y]_R$. $x \in A = [y]_R$ implies immediately that $y\,R\,x$, whence $x \in \texttt{fld}(R)$, which completes the proof of goal 2a.

2b: Let x be an arbitrarily chosen element of $\texttt{fld}(R)$: our new goal is to show that $x \in \bigcup P_R$. Since $x \in \texttt{fld}(R)$, we may choose a y such that one of $x\,R\,y$ or $y\,R\,x$ is true. But then both are true because R is symmetric, and we have $x \in [y]_R$. From $x \in [y]_R$ and $[y]_R \in P_R$, we deduce $x \in \bigcup P_R$, completing the proof of goal 2b.

Since any element of either set has been shown to belong to the other, the two sets are equal, completing the proof of Goal 2.

Goal 3: P_R **is pairwise disjoint.** Our goal is to show that for any elements A, B of P_R we have $A = B \vee A \cap B = \emptyset$ To prove this, we assume that A and B are distinct and take $A \cap B = \emptyset$ as our new goal. We prove this by contradiction: assume $A \cap B \neq \emptyset$ and our new goal is a contradiction. Since $A \cap B \neq \emptyset$, we may choose an $x \in A \cap B$. Since $A, B \in P_R$ we may choose y and z such that $A = [y]_R$ and $B = [z]_R$. If we had $y = z$ we would have $A = B$ and a contradiction, so we must have $y \neq z$. $x \in A \cap B = [y]_R \cap [z]_R$ implies $x \in [y]_R$ and $x \in [z]_R$, whence we have $x\,R\,y$ and $x\,R\,z$, whence by symmetry and transitivity of R we have $y\,R\,z$. We now prove $A = [y]_R = [z]_R = B$, which will give the desired contradiction since A and B were initially supposed distinct.

3a: Let u be an arbitrarily chosen element of $[y]_R$. Our new goal is $u \in [z]_R$. $u \in [y]_R$ implies $y\,R\,u$, and $y\,R\,z$ and symmetry imply $z\,R\,y$. Thus by transitivity of R we have $z\,R\,u$ and so $u \in [z]_R$. This completes the proof of goal 3a.

3b: Let u be an arbitrarily chosen element of $[z]_R$. Our new goal is $u \in [y]_R$. $u \in [z]_R$ implies $z\,R\,u$, which in combination

with $y\,R\,z$ and transitivity of R implies $y\,R\,u$, which implies $u \in [y]_R$, which completes the proof of goal 3b.

Since the sets $[y]_R = A$ and $[z]_R = B$ have the same elements, it follows that they are equal, which completes the proof of a contradiction, from which Goal 3 and the Theorem follow.

Theorem: If $\mathcal{P}$ is a partition of A, the relation

$$\equiv_{\mathcal{P}} = \{\langle x, y\rangle \mid (\exists B \in \mathcal{P}.x \in B \wedge y \in B)\}$$

is an equivalence relation with field A.

Proof: This is left as an exercise.

Observation: Further, $\equiv_{P_R} = R$ and $P_{\equiv_{\mathcal{P}}} = \mathcal{P}$ for any R and $\mathcal{P}$: there is a precise correspondence between equivalence relations and partitions.

An equivalence relation R represents a way in which elements of its field are similar: in some mathematical constructions we wish to *identify* objects which are similar in the way indicated by R. One way to do this is to replace references to an $x \in \mathtt{fld}(R)$ with references to its equivalence class $[x]_R$. Note that for all x, y in $\mathtt{fld}(R)$ we have $x\,R\,y$ iff $[x]_R = [y]_R$.

It might be found inconvenient that $[x]_R$ is one type higher than x. In such a situation, we would like to work with a representative of each equivalence class.

Definition: Let P be a partition. A *choice set* for P is a set C with the property that $B \cap C$ has exactly one element for each $B \in P$.

A choice set for the partition P_R will give us exactly one element of each equivalence class under R, which we can then use to represent all elements of the equivalence class in a context in which R-equivalent objects are to be identified.

In some situations, there is a natural way to choose an element of each equivalence class (a canonical representative of the class). We will see examples of this situation. In the general situation, we can invoke the last axiom of our typed theory of sets.

Axiom of Choice: If P is a partition (a pairwise disjoint set of nonempty sets) then there is a choice set C for P.

The Axiom of Choice is a somewhat controversial assertion with profound consequences in set theory: this seemed like a good place to slip it in quietly without attracting too much attention.

Here we also add some terminology about partial orders.

It is conventional when working with a particular partial order $\leq$ to use $<$ to denote $[\leq] - [=]$ (the corresponding strict partial order), $\geq$ to denote $[\leq]^{-1}$ (which is also a partial order) and $>$ to denote the strict partial order $[\geq] - [=]$.

A minimum of $\leq$ is an element m of $\mathtt{fld}(\leq)$ such that $m \leq x$ for all $x \in \mathtt{fld}(x)$. A maximum of $\leq$ is a minimum of $\geq$. A minimal element with respect to $\leq$ is an element m such that for no x is $x < m$. A maximal element with respect to $\leq$ is a minimal element with respect to $\geq$. Notice that a maximum or minimum is always unique if it exists. A minimum is always a minimal element. The converse is true for linear orders but not for partial orders in general.

For any partial order $\leq$ and $x \in \mathtt{fld}(\leq)$, we define $\mathtt{seg}_{\leq}(x)$ as $\{y \mid y < x\}$ (notice the use of the strict partial order) and $(\leq)_x$ as $[\leq] \cap (\mathtt{seg}_{\leq}(x))^2$. The first set is called the *segment* in $\leq$ determined by x and the second is called the *segment restriction* determined by x.

For any subset A of $\mathtt{fld}(\leq)$, we say that an element x of $\mathtt{fld}(\leq)$ is a lower bound for A in $\leq$ iff $x \leq a$ for all $a \in A$, and an upper bound for A in $\leq$ iff $a \leq x$ for all $a \in A$. If there is a lower bound x of A such that for every lower bound y of A, $y \leq x$, we call this the greatest lower bound of A, written $\inf_{\leq}(A)$, and if there is an upper bound x of A such that for all upper bounds y or A, we have $x \leq y$, we call this the least upper bound of A, written $\sup_{\leq}(A)$.

A special kind of partial order is a *tree*: a partial order $\leq_T$ with field T is a *tree* iff for each $x \in T$ the restriction of $\leq_T$ to $\mathtt{seg}_{\leq_T}(x)$ is a well-ordering. A subset of T which is maximal in the inclusion order among those well-ordered by $\leq_T$ is called a *branch*.

3.8.1 Exercises

1. Suppose that P is a partition.

   Prove that the relation $\sim_P$ defined by
   $x \sim_P y$ iff $(\exists A \in P. x \in A \wedge y \in A)$
   is an equivalence relation. What is the field of this equivalence relation?

Describe its equivalence classes.

This is an exercise in carefully writing everything down, so show all details of definitions and proof strategy, as far as you can.

2. This question relies on ordinary knowledge about the reals and the rationals, and also knowledge of Lebesgue measure if you have studied this (if you haven't, don't worry about that part of the question).

   Verify that the relation on real numbers defined by "$x\,R\,y$ iff $x - y$ is rational" is an equivalence relation.

   Describe the equivalence classes under this relation in general. Describe two or three specific ones. Note that each of the equivalence classes is countably infinite (why?), distinct equivalence classes are disjoint from each other, and so we "ought" to be able to choose a single element from each class.

   Can you think of a way to do this (you will not be able to find one, but thinking about why it is difficult is good for you)?

   Suppose we had a set X containing exactly one element from each equivalence class under R. For each rational number q, let X_q be the set $\{r + q \mid r \in X\}$. Note that X_q is just a translation of X.

   Prove that $\{X_q \mid q \in \mathbb{Q}\}$ is a partition of $\mathbb{R}$. (This will include a proof that the union of the X_q's is the entire real line).

   If you know anything about Lebesgue measure, you might be able to prove at this point that X is not Lebesgue measurable (if you can, do so). It is useful to note that the collection of X_q's is countable.

3.9 Cardinal Number and Arithmetic

We say that two sets are the same size iff there is a one-to-one correspondence (a bijection) between them.

Definition: We say that sets A and B are *equinumerous* and write $A \sim B$ just in case there is a bijection f from A onto B.

Theorem: Equinumerousness is an equivalence relation.

Indication of Proof: It is reflexive because the identity function on any set is a function. It is symmetric because the inverse of a bijection is a

bijection. It is transitive because the composition of two bijections is a bijection.

Definition: For any set A, we define $|A|$, the *cardinality of* A, as $[A]_\sim = \{B \mid B \sim A\}$. Notice that $|A|$ is one type higher than A. We define `Card`, the set of all *cardinal numbers*, as $\{|A| \mid A \in V\}$.

The same definitions would work if we were using the Kuratowski pair, and in fact the cardinals would be precisely the same sets.

We have already encountered some cardinal numbers.

Theorem: Each natural number is a cardinal number.

Proof: $|\emptyset| = \{\emptyset\} = 0$ is obvious: there is a bijection from $\emptyset$ to A iff $A = \emptyset$.

Suppose that $n \in \mathbb{N}$ is a cardinal number: show that $n+1$ is a cardinal number and we have completed the proof that all natural numbers are cardinals by mathematical induction. Let x be an element of n. There is a $y \notin x$ because $x \neq V$ (by the Axiom of Infinity). It suffices to show $n + 1 = |x \cup \{y\}|$. To show this, we need to show that for any set z, $z \in n+1$ iff $z \sim x \cup \{y\}$. If $z \in n+1$ then $z = v \cup \{w\}$ for some $v \in n$ and some $w \notin v$. Because n is a cardinal number there is a bijection f from x to v: $f \cup \{\langle y, w\rangle\}$ is readily seen to still be a bijection. Now let z be an arbitrarily chosen set such that $z \sim x \cup \{y\}$. This is witnessed by a bijection f. Now $f^{-1}\text{“}x$ belongs to n because n is a cardinal number, and thus we see that $v = f^{-1}\text{“}x \cup \{f^{-1}(y)\}$ belongs to $n+1$ (certainly $f^{-1}(y) \notin f^{-1}\text{“}x$), completing the proof.

There is at least one cardinal number which is not a natural number.

Definition: We define $\aleph_0$ as $|\mathbb{N}|$. Sets of this cardinality are said to be *countable* or *countably infinite*. Infinite sets not of this cardinality (if there are any) are said to be *uncountable* or *uncountably infinite*.

We provide some lemmas for construction of bijections from other bijections.

Lemma: The union of two relations is of course a relation. The union of two functions is a function iff the functions agree on the intersection of their domains: that is, if f and g are functions, $f \cup g$ is a function iff

for every $x \in \mathtt{dom}(f) \cap \mathtt{dom}(g)$ we have $f(x) = g(x)$, or, equivalently but more succinctly, $f\lceil\mathtt{dom}(f) \cap \mathtt{dom}(g) = g\lceil\mathtt{dom}(f) \cap \mathtt{dom}(g)$. Note that it is sufficient for the domains of f and g to be disjoint.

Definition: A function f is said to *cohere* with a function g iff $f\lceil\mathtt{dom}(f) \cap \mathtt{dom}(g) = g\lceil\mathtt{dom}(f) \cap \mathtt{dom}(g)$.

Lemma: The union of two injective functions f and g is an injective function iff f coheres with g and f^{-1} coheres with g^{-1}. Note that it is sufficient for the domain of f to be disjoint from the domain of g and the range of f disjoint from the range of g.

Lemma: For any x and y, $\{\langle x, y\rangle\}$ is an injection.

Arithmetic operations have natural definitions.

A cardinal $|A|$ is the collection of all sets of the same size as A. Thus, if κ is a cardinal, we mean by "set of size κ" simply an element of κ. This is not true of all representations of cardinality: if we used a representative set the same size as A as $|A|$, for example, then a set of size κ would be a set equinumerous with κ (the representation used in the usual set theory introduced later is of this latter kind).

We define addition of cardinals. Informally, a set of size $\kappa + \lambda$ will be the union of two disjoint sets, one of size κ and one of size λ.

Definition (abstract definition of addition): If κ and λ are cardinals, we define $\kappa + \lambda$ as

$$\{A \cup B \mid A \in \kappa \wedge B \in \lambda \wedge A \cap B = \emptyset\}.$$

There are some things to verify about this definition. One has to verify that $\kappa + \lambda$ is nonempty. If $A \in \kappa$ and $B \in \lambda$ then $A \times \{0\} \in \kappa$, $B \times \{1\} \in \lambda$, and these sets are obviously disjoint. The fact that cartesian product is a type level operation is crucial here (so Infinity is required). One has to verify that $\kappa + \lambda$ is a cardinal.

Observation: $|A| + |B| = |(A \times \{0\}) \cup (B \times \{1\})|$.

Proof: Suppose A' and B' are disjoint sets with bijections $f : A \to A'$ and $g : B \to B'$. Then $(\pi_1|f) \cup (\pi_1|g)$ is a bijection from $(A \times \{0\}) \cup (B \times \{1\})$ to $A' \cup B'$. The union of these two injections is an injection because they have disjoint domains and disjoint ranges, and the union has the correct domain and range.

It is perhaps preferable to simply take the Observation as the

Definition (concrete definition of addition): $|A| + |B|$ is defined as

$$|(A \times \{0\}) \cup (B \times \{1\})|.$$

(It is straightforward to show that this does not depend on the choice of representatives A and B from the cardinals).

The abstract definition of addition would work if we were using Kuratowski pairs but the proof that addition is total would be somewhat harder. The Observation would be incorrect and in fact would not make sense because it would not be well-typed.

Notice that the definition of $\kappa + 1$ as an addition of cardinals agrees with the definition of $\kappa + 1$ as a set already given in the development of finite number.

Before discussing multiplication, we consider the notion of being the same size appropriate to sets at different types.

Definition (alternative notation for singleton set): We define $\iota(x)$ as $\{x\}$. The point of this notation is that it is iterable: we can use $\iota^{\mathbf{n}}(x)$ to denote the n-fold singleton of x. [But do notice that this is not an example of iteration as ι is not a function (a function does not raise type). The n in $\iota^{\mathbf{n}}(x)$ is a purely formal bit of notation (like a type index) and not a reference to any natural number in our theory, and this is why it is in boldface]

Definition (singleton image operations): We define $\iota^{\mathbf{n}}\text{“}x$, the *$n$-fold singleton image of x* as $\{\iota^{\mathbf{n}}(y) \mid y \in x\}$. For any relation R, we define $R^{\iota^{\mathbf{n}}}$ as $\{\langle \iota^{\mathbf{n}}(x), \iota^{\mathbf{n}}(y)\rangle \mid x\, R\, y\}$. We define $T^{\mathbf{n}}(\kappa)$ for any cardinal κ as $|\iota^{\mathbf{n}}\text{“}A|$ for any $A \in \kappa$. Note that $A \sim B \leftrightarrow \iota\text{“}A \sim \iota\text{“}B$ is obvious: if f is a bijection from A to B, then f^{ι} will be a bijection from $\iota\text{“}A$ to $\iota\text{“}B$. We define $T^{-\mathbf{n}}(\kappa)$ as the unique cardinal λ (if there is one) such that $T^{\mathbf{n}}(\lambda) = \kappa$.

Definition (sole element): We define $\iota^{-\mathbf{1}}(\{x\})$ as x. We define $\iota^{-\mathbf{1}}(A)$ as $\emptyset$ if A is not a singleton. $\iota^{-\mathbf{n}}(\iota^{\mathbf{n}}(x))$ will be defined as x as one might expect, if this notation is ever needed.

The singleton map (or iterated singleton map) is in a suitable external sense injective, so a set equinumerous with $\iota^{\mathbf{n}}$"A, though it is n types higher than A, is in a recognizable sense the same size as A.

The definition of $T^{-\mathbf{n}}$ depends on the observation that $T^{\mathbf{n}}$ is an externally injective map from cardinals in type i to cardinals in type $i+n$, so if there is a suitable λ there is only one. We leave open the possibility that $T^{-\mathbf{n}}(\kappa)$ is undefined for some cardinals κ and indeed this turns out to be the case.

We discuss the application of the T operation to natural numbers.

Theorem: $T(0) = 0$ and $T(n+1) = T(n) + 1$.

Corollary: $T(1) = 1; T(2) = 2; T(3) = 3 \ldots$ But we cannot say

$$(\forall n \in \mathbb{N}.T(n) = n),$$

because this is ungrammatical.

Theorem: For all natural numbers n, $T(n)$ is a natural number. For all natural numbers n [not of the lowest type] $T^{-\mathbf{1}}(n)$ exists and is a natural number.

Proof: We prove both parts by induction, of course.

Our first goal is to prove that $T(n)$ is a natural number for every natural number n. We observe first that $T(0) = 0$ is obvious, as ι"$\emptyset = \emptyset$. Now suppose that k is a natural number and $T(k)$ is a natural number. Our aim is to prove that $T(k+1)$ is a natural number. Each element of $k+1$ is of the form $A \cup \{x\}$ where $A \in k$ and $x \notin A$. $T(k+1) = |\iota"(A \cup \{x\})|$. But $\iota"(A \cup \{x\}) = \iota"A \cup \{\{x\}\}$. Obviously $\iota"A \in T(k)$ and $\{x\} \notin \iota"A$, so $\iota"A \cup \{\{x\}\} \in T(k) + 1 \in \mathbb{N}$, so $T(k+1) = T(k) + 1 \in \mathbb{N}$.

Our second goal is to prove that $T^{-\mathbf{1}}(n)$ exists and is a natural number for each natural number n (not of the lowest possible type). Since $T(0) = 0$, we also have $T^{-\mathbf{1}}(0) = 0$, so $T^{-\mathbf{1}}(0)$ exists and is a natural number. Let k be a natural number such that there is a natural number l such that $T(l) = k$ (which is equivalent to saying that $T^{-\mathbf{1}}(k)$ exists and is a natural number). Choose a set A of cardinality l. Choose $x \notin A$. $|A \cup \{x\}| = l + 1$ and $|\iota"(A \cup \{x\})| = |\iota"A \cup \{\{x\}\}| = k + 1$ is obvious, so $T(l+1) = k+1$, whence $T^{-\mathbf{1}}(k+1)$ exists and is a natural number as desired.

Reasonable Convention: It is reasonable to simply identify the natural numbers at different types and there is a way to make sense of this in our notation: allow a natural number variable n of type k to appear at other types with the understanding that where it appears in a position appropriate for a variable of type $k+i$ it is actually to be read as $T^i(n)$. We will not do this, or at least we will explicitly note use of this convention if we do use it, but it is useful to note that it is possible.

Rosser's Counting Theorem: $\{1,\ldots,n\} \in T^{\mathbf{2}}(n)$, for each positive natural number n.

Discussion and Proof: Of course $\{1,\ldots,n\} = \{m \in \mathbb{N} \mid 1 \leq m \leq n\}$ has n members, if n is a concrete natural number. But the second n we mention is two types higher than the first one: we fix this by affixing $T^{\mathbf{2}}$ to the second one, so that both occurrences of n have the same type.

What this actually says is that if we have a set A belonging to a natural number n, we can put $\iota^{\mathbf{2}}{}``A$ (the set of double singletons of elements of A) into one-to-one correspondence with the set of natural numbers $\{1,\ldots,n\}$ of the type appropriate for $A \in n$ to make sense. This can be proved by induction on the number of elements in A. If A has one element a, clearly there is a bijection between $\{\{\{a\}\}\}$ and $\{1\}$ (all that needs to be checked is that these objects are of the same type: the number 1 being considered satisfies $A \in 1$). Suppose that for all $A \in n$, $\iota^{\mathbf{2}}{}``A \sim \{1,\ldots,n\}$. We want to show that for any $B \in n+1$, $\iota``B \sim \{1,\ldots,n+1\}$. $B = A \cup \{x\}$ for some $A \in n, x \notin A$. There is a bijection f from $\iota^{\mathbf{2}}{}``A$ to $\{1,\ldots,n\}$ by inductive hypothesis. $f \cup \langle\{\{x\}\}, n+1\rangle$ is easily seen to witness the desired equivalence in size.

Von Neumann's Counting Theorem: For any natural number n,

$$\{m \in \mathbb{N} \mid m < n\} \in T^{\mathbf{2}}(n).$$

Discussion: This is true for the same reasons. It is not really a theorem of von Neumann, but it relates to his representation of the natural numbers.

Notice that these counting theorems could be written in entirely unexciting forms if we adopted the Reasonable Convention above. It would then

be the responsibility of the reader to spot the type difference and insert the appropriate T operation. This would have to be done in order to *prove* either of these statements.

A fully abstract definition of multiplication would say that $\kappa \cdot \lambda$ is the size of the union of κ disjoint sets each of size λ. To state this precisely requires the T operation just introduced.

*__Definition (abstract definition of multiplication):__ $\kappa \cdot \lambda$ is the uniquely determined cardinal of a set $\bigcup C$ where C is pairwise disjoint, $C \in T(\kappa)$, and $C \subseteq \lambda$.

The details of making this definition work are quite laborious. Infinity is required to show that there are such sets for any κ and λ, and Choice is required to show that the cardinal is uniquely determined. We regretfully eschew this definition and use a more concrete definition employing the cartesian product:

Definition (concrete definition of multiplication): $|A| \cdot |B|$ is defined as $|A \times B|$. It is straightforward to show that this does not depend on the choice of representatives A, B from the cardinals.

If we were using the Kuratowski pair we would define

$$|A| \cdot |B| = T^{-\mathbf{2}}(|A \times B|).$$

It would be harder to show that multiplication is total. We would also have

$$|A| + |B| = T^{-\mathbf{2}}(|(A \times \{0\}) \cup (B \times \{1\})|)$$

if we were using the Kuratowski pair.

The T operation commutes with arithmetic operations:

Theorem: For all cardinal numbers κ and λ, $T(\kappa) + T(\lambda) = T(\kappa + \lambda)$ and $T(\kappa \cdot \lambda) = T(\kappa) \cdot T(\lambda)$.

Theorems of cardinal arithmetic familiar from the theory of natural numbers (and from ordinary experience) have much more natural proofs in set theory than the inductive proofs given in Peano arithmetic.

Theorem: The following identities are true for all cardinal numbers κ, λ, μ (including natural numbers).

1. $\kappa + 0 = \kappa; \kappa \cdot 1 = \kappa$
2. $\kappa \cdot 0 = 0$
3. $\kappa + \lambda = \lambda + \kappa; \kappa \cdot \lambda = \lambda \cdot \kappa$
4. $(\kappa + \lambda) + \mu = \kappa + (\lambda + \mu); (\kappa \cdot \lambda) \cdot \mu = \kappa \cdot (\lambda \cdot \mu)$
5. $\kappa \cdot (\lambda + \mu) = \kappa \cdot \lambda + \kappa \cdot \mu$

All of these admit very natural proofs.

Sample Proofs:

commutativity of multiplication: Let κ, λ be cardinal numbers. Choose sets A and B such that $\kappa = |A|$ and $\lambda = |B|$. $\kappa \cdot \lambda = |A \times B|$ and $\lambda \cdot \kappa = |B \times A|$; what remains is to show that there is a bijection from $|A \times B|$ to $|B \times A|$. The map which sends each ordered pair $\langle a, b\rangle$ (for $a \in A, b \in B$) to $\langle b, a\rangle$ does the trick.

associativity of addition: Let κ, λ, μ be cardinal numbers. Choose A, B, C such that $\kappa = |A|, \lambda = |B|, \mu = |C|$. $(|A| + |B|) + |C| = |A \times \{0\} \cup B \times \{1\}| + |C| = |(A \times \{0\} \cup B \times \{1\}) \times \{0\} \cup C \times \{1\}| = |\{\langle\langle a, 0\rangle, 0\rangle \mid a \in A\} \cup \{\langle\langle b, 1\rangle, 0\rangle \mid b \in B\} \cup \{\langle c, 1\rangle \mid c \in C\}|$

Similarly $|A| + (|B| + |C|) = |\{\langle a, 0\rangle \mid a \in A\} \cup \{\langle\langle b, 0\rangle, 1\rangle \mid b \in B\} \cup \{\langle\langle c, 1\rangle, 1\rangle \mid c \in C\}|$.

A bijection from $\{\langle\langle a, 0\rangle, 0\rangle \mid a \in A\} \cup \{\langle\langle b, 1\rangle, 0\rangle \mid b \in B\} \cup \{\langle c, 1\rangle \mid c \in C\}$ to $\{\langle a, 0\rangle \mid a \in A\} \cup \{\langle\langle b, 0\rangle, 1\rangle \mid b \in B\} \cup \{\langle\langle c, 1\rangle, 1\rangle \mid c \in C\}$ is provided by the union of the map sending each $\langle\langle a, 0\rangle, 0\rangle$ to $\langle a, 0\rangle$, the map sending each $\langle\langle b, 1\rangle, 0\rangle$ to $\langle\langle b, 0\rangle, 1\rangle$ and the map sending each $\langle c, 1\rangle$ to $\langle\langle c, 1\rangle, 1\rangle$. Each of these maps is a bijection, they have disjoint domains and disjoint ranges, so their union is still a bijection. The existence of this bijection witnesses the desired equation.

Important arithmetic properties of the natural numbers *not* shared by general cardinals are the *cancellation properties*. It is not true in general that $\kappa + \mu = \lambda + \mu \leftrightarrow \kappa + \lambda$, nor that $\kappa \cdot \mu = \lambda \cdot \mu \wedge \mu \neq 0 \rightarrow \kappa = \lambda$. This means that we do not get sensible notions of subtraction or division.

But the following is a

Theorem: For any cardinals κ, λ and any natural number n, $\kappa + n = \lambda + n \rightarrow \kappa = \lambda$.

Proof: It suffices to prove $\kappa + 1 = \lambda + 1 \to \kappa = \lambda$: the result then follows by induction.

Suppose $\kappa+1 = \lambda+1$. Let A and B be chosen so that $\kappa = |A|$, $\lambda = |B|$, and neither A nor B is the universal set V. Note that if either A or B were the universal set, we could replace it with $V \times \{0\} \sim V$. Choose $x \notin A, y \notin B$. We have $|A \cup \{x\}| = \kappa+1 = \lambda+1 = |B \cup \{y\}|$. This means we can choose a bijection $f : (A \times \{x\}) \to (B \times \{y\})$. Either $f(x) = y$ or $f(x) \neq y$. If $f(x) = y$, then $f \restriction A$ is the desired bijection from A to B, witnessing $\kappa = \lambda$. If $f(x) \neq y$, then $f - \{\langle x, f(x)\rangle\} - \{\langle f^{-1}(y), y\rangle\} \cup \{\langle f^{-1}(y), f(x)\rangle$ is the desired bijection from A to B witnessing $\kappa = \lambda$. In either case we have established the desired conclusion.

3.9.1 Exercises

1. Prove that $|\mathbb{N}| + 1 = |\mathbb{N}| + |\mathbb{N}| = |\mathbb{N}| \cdot |\mathbb{N}| = |\mathbb{N}|$.

   Describe bijections by arithmetic formulas where you can; in any case clearly describe how to construct them (these are all familiar results, or should be, and all of the bijections can in fact be described algebraically: the formula for triangular numbers can be handy for this). I'm looking for bijections with domain $\mathbb{N}$ and range some more complicated set in every case.

2. Verify the distributive law of multiplication over addition in cardinal arithmetic,

   $$|A| \cdot (|B| + |C|) = |A| \cdot |B| + |A| \cdot |C|,$$

   by writing out explicit sets with the two cardinalities (fun with cartesian products and labelled disjoint unions!) and explicitly describing the bijection sending one set to the other. You do not need to prove that it is a bijection: just describe the sets and the bijection between them precisely.

3. Prove that $|\langle A, B\rangle| = |A| + |B|$ if the pair is taken to be a Quine pair.

4. Explain why the relation $A \sim B$ of equinumerousness (equipotence, being the same size) is an equivalence relation by citing basic properties of bijections.

   The structure of your proof should make it clear that you understand what an equivalence relation is.

You do not need to prove the basic properties of bijections that are needed; you need only state them.

Your proof should also make it clear that you know what $A \sim B$ means.

What are the equivalence classes under the relation $\sim$ called in type theory?

5. In this problem you will indicate a proof of the associative property of multiplication for cardinal numbers.

   Recall that $|A| \cdot |B|$ is defined as $|A \times B|$.

   The goal is to prove that $(|A| \cdot |B|) \cdot |C| = |A| \cdot (|B| \cdot |C|)$. Describe sets of these cardinalities and (carefully) describe a bijection between them. You do not need to prove that the map is a bijection.

3.10 Number Systems

In this section we give a development of the system of real numbers from the typed theory of sets. Part of the point is that this development is not unique or canonical in any way: we indicate how alternative developments might go. The development is full in the sense that all definitions of mathematical structures are given. Not all theorems are proved, though important ones are stated.

We begin with the system $\mathbb{N}^+$ of all nonzero natural numbers. We have already defined arithmetic operations of addition and multiplication on the natural numbers, and it is easy to see that $\mathbb{N}^+$ is closed under these operations.

We now give a construction of the system $\mathbb{Q}^+$ of *fractions* (positive rational numbers).

Definition: For $m, n \in \mathbb{N}^+$, we define $m|n$ as $(\exists x \in \mathbb{N}^+ . m \cdot x = n)$. This is read "$n$ is divisible by m" and we say that m is a *factor* of n.

Definition: For $m, n \in \mathbb{N}^+$, we define $\texttt{gcd}(m, n)$ as the largest natural number x which is a factor of m and a factor of n. If $\texttt{gcd}(m, n) = 1$, we say that m and n are *relatively prime*.

Theorem: If $m \cdot x = m \cdot y$, then $x = y$, where $m, x, y \in \mathbb{N}^+$.

Definition: If $m \cdot x = n$, we define $\frac{n}{m}$ as x (this is uniquely determined, if defined, by the previous theorem). Note that this notation will be superseded after the following definition.

Definition: We define a *fraction* as an ordered pair $\langle m, n \rangle$ of nonzero natural numbers such that m and n are relatively prime. For any ordered pair $\langle m, n \rangle$ of nonzero natural numbers, we define $\texttt{simplify}(m, n)$ as $\left\langle \frac{m}{\texttt{gcd}(m,n)}, \frac{n}{\texttt{gcd}(m,n)} \right\rangle$. Note that $\texttt{simplify}(m, n)$ is a fraction. After this point, we use the notation $\frac{m}{n}$ to denote $\texttt{simplify}(m, n)$.

Observation: It is more usual to define an equivalence relation $\langle m, n \rangle \sim \langle p, q \rangle$ on ordered pairs of nonzero natural numbers (usually actually ordered pairs of integers with nonzero second projection) as holding when $mq = np$ (a proof that this is an equivalence relation is needed) then define fractions (more usually general rationals) as equivalence classes under this relation. The construction given here uses canonical representatives instead of equivalence classes.

Definition: We define $\frac{m}{n} + \frac{p}{q}$ as $\frac{mq+np}{pq}$ and $\frac{m}{n} \cdot \frac{p}{q} = \frac{mp}{nq}$. We define $\frac{m}{n} \leq \frac{p}{q}$ as holding iff $mq \leq np$. We leave it to the reader to prove that these definitions are valid (do not depend on the choice of representation for the fractions), that $\leq$ is a linear order, and that addition and multiplication of fractions have expected properties. The complete familiarity of these definitions may obscure the fact that work needs to be done here.

Now we proceed to define the system of *magnitudes* (positive real numbers).

Definition: A *magnitude* is a set m of fractions with the following properties.

1. m and $\mathbb{Q}^+ - m$ are nonempty.
2. $(\forall pq \in \mathbb{Q}^+ . p \in m \wedge q \leq p \rightarrow q \in m)$: m is downward closed.
3. $(\forall p \in m . (\exists q \in m . p \leq q))$: m has no largest element.

The motivation here is that for any positive real number r (as usually understood prior to set theory) the intersection of the interval $(0, r)$ with

the set of positive rationals uniquely determines r (and of course is uniquely determined by r) and any set of positive rationals m with the properties given above will turn out to be the intersection of the set of positive rationals and $(0, \sup m)$.

Definition: For magnitudes m and n, we define $m + n$ as

$$\{p + q \mid p \in m \wedge q \in n\}$$

and $m \cdot n$ as

$$\{p \cdot q \mid p \in m \wedge q \in n\}.$$

We define $m \leq n$ as $m \subseteq n$. We leave it to the reader to prove that addition and multiplication of magnitudes always yield magnitudes and that these operations and the order relation have the expected properties.

This is where the payoff of our particular approach is found. It is more usual to use intersections of intervals $(-\infty, r)$ with all the reals to represent the reals; with this representation of reals the definition of multiplication is horrible.

We cite a

Theorem: If $m + x = m + y$ then $x = y$, for magnitudes m, x, y.

Definition: If $m + x = n$, we define $n - m$ as x (uniqueness of $n - m$ if it exists follows from the previous theorem). This definition will be superseded by the following definition.

Definition: We define a *real number* as an ordered pair of magnitudes one of which is equal to 1 (where the magnitude 1 is the set of all fractions less than the fraction $1 = \frac{1}{1}$). For any pair of magnitudes $\langle x, y \rangle$, we define $\texttt{simp}(x, y)$ as $\langle (x + 1) - \min(x, y), (y + 1) - \min(x, y) \rangle$. Notice that $\texttt{simp}(x, y)$ will be a real number. Denote $\texttt{simp}(x, y)$ by $x - y$ (superseding the previous definition).

Definition: We define $(x - y) + (u - v)$ as $(x + u) - (y + v)$. We define $(x-y)\cdot(u-v)$ as $(xu+yv)-(xv+yu)$. We define $x-y \leq u-v$ as holding precisely when $x+v \leq y+u$. We leave it to the reader to establish that everything here is independent of the specific representation of $x - y$ and $u - v$ used, and that the operations and the order relation have expected properties.

A considerable amount of overloading is found here. Addition, multiplication and order are already defined for nonzero natural numbers when we start. In each system, addition, multiplication, and order are defined: these are different operations and relations in each system. Names of nonzero natural numbers, fractions, and magnitudes are also overloaded: the natural number n is confused with the fraction $\frac{n}{1}$ but it is not the same object, and similarly the magnitude $\{q \in \mathbb{Q}^+ \mid q < p\}$ is not the same object as the fraction p, and the real number $(m + 1) - 1$ is not the same object as the magnitude m, though in each case we systematically confuse them.

Certain important subsystems do not have a place in our development though they do in more usual developments.

Definition: We define the real number 0 as $1 - 1$. For each real number $r = x - y$ we define $-r$ as $y - x$. We define $r - s$ as $r + (-s)$ for reals r and s.

Definition: We define the set of *integers* $\mathbb{Z}$ as the union of the set of all (real numbers identified with) nonzero naturals, $\{0\}$, and the set of all additive inverses $-n$ of (real numbers identified with) nonzero naturals n.

Definition: We define the set of *rationals* $\mathbb{Q}$ as the union of the set of all (real numbers identified with) fractions p, $\{0\}$, and the set of all additive inverses $-p$ of (real numbers identified with) fractions p.

Definition: For any fraction $q = \frac{m}{n}$ we define q^{-1} as $\frac{n}{m}$. For any magnitude m, we define m^{-1} as $\{q^{-1} \mid q \notin m\}$. It is straightforward to prove that m^{-1} is a magnitude and $m \cdot m^{-1} = 1$ for each m. Now define the reciprocal operation for reals: $((m + 1) - 1)^{-1} = (m^{-1} + 1) - 1$ and $(1 - (m + 1))^{-1} = 1 - (m^{-1} + 1)$ for each magnitude m, while $(1 - 1)^{-1}$ is undefined. It can be proved that $r \cdot r^{-1} = 1$ for each real $r \neq 0$. Finally, we define $\frac{r}{s}$ as $r \cdot s^{-1}$ for any real r and nonzero real s.

We noted above that we have avoided the use of equivalence classes of ordered pairs at the steps passing to fractions and to signed real numbers, preferring to use canonical representatives. Simplification of fractions is of course a familiar mathematical idea; the canonical representation of reals we use is less obvious but works just as well.

In this development we have followed the prejudices of the ancient Greeks as far as possible, delaying the introduction of zero or negative quantities to the last step.

The reals as defined here satisfy the following familiar axioms of a "complete ordered field". Up to a suitable notion of isomorphism, the reals are the only complete ordered field.

commutative laws: $a + b = b + a$; $a \cdot b = b \cdot a$.

associative laws: $(a + b) + c = a + (b + c)$; $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

distributive law: $a \cdot (b + c) = a \cdot b + a \cdot c$.

identity laws: $a + 0 = a$; $a \cdot 1 = a$.

inverse laws: $a + (-a) = 0$; $a \cdot a^{-1} = 1$ if $a \neq 0$.

nontriviality: $0 \neq 1$

closure of positive numbers: If $a \geq 0$ and $b \geq 0$ then $a + b \geq 0$ and $a \cdot b \geq 0$. [note that $a \geq 0$ is a primitive notion at this point in the development: the reals of the form $r = (m + 1) - 1$ are the ones for which we assert $r \geq 0$].

trichotomy: For each real number a, exactly one of the following is true: $a \geq 0$, $a = 0$, $-a \geq 0$.

Definition: $a \leq b$ iff $b + (-a) \geq 0$.

Theorem: $\leq$ thus defined is a linear order.

completeness: Any nonempty set of reals which is bounded above (in terms of the order just defined) has a least upper bound.

3.11 Well-Orderings and Ordinal Numbers

We recall that a well-ordering is a linear order with the property that the corresponding strict partial order is well-founded.

Definition: A *well-ordering* is a linear order $\leq$ with the property that for each nonempty subset A of $\mathtt{fld}(\leq)$ there is $a \in A$ such that there is no $b \neq a$ in A such that $b \leq a$: such an a is a minimal element of A (in fact, the minimal element is unique because $\leq$ is linear).

In this section, we study the structure of well-orderings. In this section we state and prove powerful and highly abstract theorems: for some concrete discussion of ordinal numbers, look toward the end of the next section.

Definition: Two relations R and S are said to be *isomorphic* iff there is a bijection f from $\texttt{fld}(R)$ to $\texttt{fld}(S)$ such that for all x, y, $x\,R\,y \leftrightarrow f(x)\,S\,f(y)$. f is said to be an *isomorphism* from R to S. We write $R \approx S$ for "R is isomorphic to S".

Theorem: Isomorphism is an equivalence relation on relations.

Definition: An equivalence class under isomorphism is called an *isomorphism type*.

Theorem: A relation isomorphic to a well-ordering is a well-ordering.

Definition: The isomorphism type of a well-ordering is called its *order type*. We write $\texttt{ot}(\leq)$ for the order type $[\leq]_\approx$ of $\leq$. A set is an *ordinal number* iff it is the order type of some well-ordering. The set of all ordinal numbers is called `Ord`.

There are few well-orderings familiar to us from undergraduate mathematics. Any finite linear order is a well-ordering.

Theorem: For any $n \in \mathbb{N}$, any two linear orders with field of size n are isomorphic and are well-orderings.

Theorem: A well-ordering is finite iff its converse is also a well-ordering.

Our use of "finite" in the previous theorem might cause confusion, which will be alleviated by considering the following

Lemma: A relation (considered as a set) is finite iff its field is finite.

Definition (finite ordinals): For each natural number n, there is a unique ordinal number which is the order type of all orders with range of that cardinality: we also write this ordinal number as n, though it is not the same object as the cardinal number n.

An amusing observation, depending crucially on the exact details of our implementation, is the following relationship between ordinal and cardinal numbers.

Theorem: The ordinal number n is a subset of the cardinal number $\frac{n(n-1)}{2}$.

In the usual untyped set theory, with the usual implementations of the notions of ordinal and cardinal number, the finite cardinals and the finite ordinals are the same objects. We will see this in section 4.

The usual order on the natural numbers is a well-ordering. The usual orders on the integers, rationals and reals are *not* well-orderings. Another example of an infinite well-ordering which is familiar from calculus is the order on reals restricted to the range of a strictly increasing bounded sequence taken together with its limit.

Definition: We define ω as the order type of the natural order on the natural numbers.

We give some basic definitions for arithmetic of ordinal numbers.

Definition (ordinal addition): For well-orderings R and S, we define another well-ordering $R \oplus S$. The field of $R \oplus S$ is $\mathtt{fld}(R) \times \{0\} \cup \mathtt{fld}(S) \times \{1\}$. $\langle x, i\rangle\, (R \oplus S)\, \langle y, j\rangle$ is defined as $i < j \vee i = 0 \wedge j = 0 \wedge x\, R\, y \vee i = 1 \wedge j = 1 \wedge x\, S\, y$. Intuitively, we make disjoint orders of types R and S and put the order of type R in front of the order of type S. Finally, we define $\alpha + \beta$ for ordinals α and β as $\mathtt{ot}(R \oplus S)$ for any $R \in \alpha$ and $S \in \beta$.

Another way to put this: for any relation R, define R_x as $\{\langle \langle a, x\rangle, \langle b \in x\rangle\rangle \mid a\, R\, b\}$. Notice that $R \approx R_x$ for any R and x and $R_x \cap S_y = \emptyset$ for any R and S and any distinct x and y. For any ordinals α, β define $\alpha + \beta$ as $\mathtt{ot}(R_0 \cup (\mathtt{fld}(R_0) \times \mathtt{fld}(S_1)) \cup S_1)$ where $R \in \alpha$ and $S \in \beta$. It is straightforward to establish that $R_0 \cup (\mathtt{fld}(R_0) \times \mathtt{fld}(S_1)) \cup S_1$ is a well-ordering and that its order type does not depend on which representatives R and S are chosen from α and β.

Discussion: An order of type $\omega + 1$ is readily obtained: define $x \leq' y$ as

$$x \in \mathbb{N} \wedge y \in \mathbb{N} \wedge 0 < x \leq y \vee y = 0.$$

In effect, we move 0 from its position at the beginning of the order to the end. This is the same order type as that of a strictly increasing sequence taken together with its limit, which we mentioned above.

The relation $\leq'$ is not isomorphic to the usual $\leq$ on the natural numbers. An easy way to see this is that there is a $\leq'$-largest element of the field of $\leq'$, and this is a property of relations which is preserved by isomorphism: if $\leq'\approx\leq$ were witnessed by an isomorphism f then $f(0)$ would have to be the $\leq$-largest natural number, and there is no such natural number.

Further, the field of $\leq'$ is the same size as the field of $\leq$ (in fact, it is the same set!): so the theorem that there is a unique order type of well-orderings of each finite cardinality n does not generalize to infinite cardinalities.

Observe that an order of type $\omega+\omega$ is a still more complex well-ordering with field the same size as the field of a relation of type ω. A concrete example of such an order would be the order

$$\{\langle m, n\rangle \in \mathbb{N}^2 \mid 2|(x-y) \wedge x \leq y \vee 2 \not| x \wedge 2|y\},$$

which puts the odd and even numbers in their usual respective orders but puts all the odd numbers before all the even numbers.

Definition (ordinal multiplication): For well-orderings R and S, we define another well-ordering $R \otimes S$. The field of $R \otimes S$ is $\mathtt{fld}(R) \times \mathtt{fld}(S)$. $\langle x, y\rangle\,(R \otimes S)\,\langle u, v\rangle$ is defined as $u\,S\,v \vee u = v \wedge x\,R\,y$. This is reverse lexicographic order on the cartesian product of the fields of the relations. Finally, we define $\alpha \cdot \beta$ for ordinals α and β as $\mathtt{ot}(R \otimes S)$ for any $R \in \alpha$ and $S \in \beta$.

The order $\omega \cdot \omega$ is a still more complex order type whose field is the same size as that of any relation of order type ω. There are very complicated well-orderings with countable fields (whose order types are called *countable ordinals*).

The algebra of ordinal numbers contains surprises. Some algebraic laws do work much as expected, but some basic laws are not inherited from the algebra of natural numbers. For example, $\omega + 1 \neq 1 + \omega = \omega$ and $\omega \cdot 2 \neq 2 \cdot \omega = \omega$.

We now study the natural order relation on the ordinal numbers, which turns out to be a well-ordering itself (at a higher type).

Definition: If $\leq$ is a partial order and $x \in \mathtt{fld}(\leq)$, we define $\mathtt{seg}_{\leq}(x)$ as $\{y \mid y < x\}$ (where $<$ is the strict partial order $[\leq] - [=]$). $\mathtt{seg}_{\leq}(x)$ is

the *segment* determined by x. We define $\leq_x$ as $[\leq] \cap \mathtt{seg}_{\leq}(x)^2$; this is the *segment restriction* of $\leq$ determined by x.

Theorem: If $\leq$ is a well-ordering and $x \in \mathtt{fld}(\leq)$ then $\leq_x$ is a well-ordering.

Lemma: No well-ordering is isomorphic to one of its own segment restrictions.

Proof: Suppose that $\leq$ is an isomorphism, x is in the field of $\leq$, and $\leq \approx (\leq)_x$ is witnessed by an isomorphism f. Since $f(x) \neq x$ is obvious (x is not in the range of f!), there must be a $\leq$-least y such that $f(y) \neq y$. Let $A = \mathtt{seg}_{\leq}(y)$. Each element of A is fixed by f. In $\leq$, y is the least object greater than all elements of A. In $(\leq)_x$, $f(y)$ is the least object greater than all elements of A. The two orders agree on the common part of their field. Since $f(y)$ is certainly in the field of $\leq$, we have $y \leq f(y)$ (as otherwise $f(y)$ would be a smaller strict upper bound for A in $\leq$). Since $y \leq f(y)$, we have y in the field of $(\leq)_x$, and $f(y) \leq y$, as otherwise y would be a smaller strict upper bound for A in $(\leq)_x$. So $y = f(y)$, which is a contradiction.

Corollary: No two distinct segment restrictions of the same well-ordering can be isomorphic to one another.

Proof: One of them would be a segment restriction of the other.

Definition: We say that a subset D of the field of a well-ordering $\leq$ is "downward closed in $\leq$" iff $(\forall d \in D.(\forall e \leq d.e \in D))$.

Lemma: For any well-ordering $\leq$, a set downward closed in $\leq$ is either the field of $\leq$ or a segment in $\leq$.

Proof: Let D be a set downward closed in $\leq$. If x belongs to the field of $\leq$ but does not belong to D, then $d < x$ must be true for all $d \in D$, as otherwise we would have $x \leq d \in D$, from which $x \in D$ would follow. This means that if D has no strict upper bound, it must be the entire field of $\leq$. If D does have a strict upper bound, it must have a $\leq$-least strict upper bound x because $\leq$ is a well-ordering. We claim that $D = \mathtt{seg}_{\leq}(x)$ in this case. If $y \in \mathtt{seg}_{\leq}(x)$, then y cannot be a strict upper bound of D because x is the least strict upper bound of D, and so y must be an element of D. If y is an element of D, then y

must be less than x because x is a strict upper bound of D, that is, y is an element of $\mathtt{seg}_{\leq}(x)$. Sets with the same elements are the same.

Theorem: If $\leq_1$ and $\leq_2$ are well-orderings, then exactly one of three things is true: either $\leq_1$ and $\leq_2$ are isomorphic, or $\leq_1$ is isomorphic to a segment restriction $(\leq_2)_x$, or $\leq_2$ is isomorphic to a segment restriction $(\leq_1)_x$.

Proof: Let $\leq_1$ be a well-ordering with field A. Let $\leq_2$ be a well-ordering with field B. Define C as $\{a \in A \mid \neg(\exists b \in B.(\leq_1)_a \approx (\leq_2)_b)\}$, the set of all elements of the field of $\leq_1$ whose segment restrictions are *not* isomorphic to a segment restriction in $\leq_2$. If C is nonempty, it has a least element c. Each $d <_1 c$ does not belong to C, because c is the $\leq_1$-least element of C. Thus, by the definition of C, there is an $e \in B$ such that $(\leq_1)_d \approx (\leq_2)_e$. There can be only one such e because no two segment restrictions of the same well-ordering can be isomorphic to each other. Thus there is a function F which maps each $d <_1 c$ to the unique e such that $(\leq_1)_d \approx (\leq_2)_e$. We claim that F is an isomorphism from $(\leq_1)_c$ to $\leq_2$. This breaks down into three subclaims: F is an injection, F is order-preserving, and the range of F is B. For each $d <_1 c$, we have an isomorphism f witnessing $(\leq_1)_d \approx (\leq_2)_{F(d)}$. For each $d' < d$, the restriction of f to $\mathtt{seg}_{\leq_1}(d')$ is an isomorphism from $(\leq_1)_{d'}$ to $(\leq_2)_{f(d')}$, so in fact $F(d') = f(d')$. Because the range of f is the segment in $\leq_2$ determined by $F(d)$, we have $F(d') = f(d') < F(d)$. This shows both that F is order preserving and that it is a bijection. Further, it shows that the range of F is downward closed, as we see that the restriction of F to the segment determined by d is the isomorphism from the segment determined by d to the segment determined by $F(d)$. Since the range of F is downward closed, it must be either B or some $\mathtt{seg}_{\leq_2}(x)$, so F is either an isomorphism from $(\leq_1)_c$ to $\leq_2$ or an isomorphism from $(\leq_1)_c$ to some $(\leq_2)_x$. The latter case is impossible by the definition of c, so we must actually have F an isomorphism from $(\leq)_c$ to $\leq_2$, establishing the Theorem in this case. If the set C is empty, then for every $a \in A$ there is $b \in B$ such that $(\leq_1)_a \approx (\leq_2)_b$. This b must be unique as no two distinct segment restrictions of $\leq_2$ can be isomorphic. For each $a \in A$, we define $F(a)$ as the unique b such that $(\leq_1)_a \approx (\leq_2)_b$. Exactly the same argument just given shows that F is a bijection, order-preserving, and has a downward closed range. From this it follows just

as in the first case that F is an isomorphism from $\leq_1$ to either $\leq_2$ or some $(\leq_2)_x$, establishing that the Theorem is true in this case. If $\leq_1 \approx \leq_2$ then we cannot have either $(\leq_1)_x \approx \leq_2$ or $(\leq_2)_x \approx \leq_1$ because a well-ordering cannot be similar to one of its segment restrictions. If we had $\leq_1 \approx (\leq_2)_x$, and further had $\leq_2 \approx (\leq_1)_y$, witnessed by an isomorphism g, then we would have $\leq_1 \approx (\leq_1)_{g(x)}$, which is impossible. This establishes that only one of the three cases can hold.

Definition: If α and β are ordinal numbers, we define $\alpha \leq \beta$ as holding iff either $\alpha = \beta$ or each element of α is isomorphic to a segment restriction in each element of β.

Theorem: The relation $\leq$ defined on ordinal numbers in the previous definition is a well-ordering. Where it is necessary to distinguish it from other orders, we write it $\leq_\Omega$. $\mathtt{ot}(\leq_\Omega)$ is called Ω: notice that Ω is not of the same type as the ordinals in the field of the relation $\leq_\Omega$ of which it is the order type (it is 2 types higher; it would be 4 types higher if we defined well-orderings using the Kuratowski pair).

Proof: Let α and β be ordinals. If $\leq_1 \in A$ and $\leq_2 \in B$, then either $\leq_1 \approx \leq_2$, in which case $\alpha = \beta$, or $\leq_1$ is isomorphic to a segment restriction in $\leq_2$, in which case the same is true for any $\leq_1' \approx \leq_1$ and $\leq_2' \approx \leq_2$, or $\leq_2$ is isomorphic to a segment restriction in $\leq_1$, in which case the same is true for any $\leq_2' \approx \leq_2$ and $\leq_1' \approx \leq_1$. If more than one of these alternatives held for any pair of well-orderings, one of them could be shown to be isomorphic to one of its own segment restrictions. Certainly $\alpha \leq \alpha$, so the $\leq$ relation on ordinals is reflexive. If $\alpha \leq \beta$ and $\beta \leq \alpha$ this must be witnessed by isomorphisms between $\leq_1 \in \alpha$ and $\leq_2 \in \beta$ in both directions, or once again we would have one of these well-orderings isomorphic to a segment restriction of itself. So the $\leq$ relation on ordinals is antisymmetric. If we have $\alpha \leq \beta$ and $\beta \leq \gamma$ and we choose $\leq_1, \leq_2, \leq_3$ in α, β, γ respectively, we have $\leq_1$ isomorphic to $\leq_2$ or a segment restriction thereof, and $\leq_2$ isomorphic to $\leq_3$ or a segment restriction thereof, and composition of isomorphisms gives us an isomorphism from $\leq_1$ to $\leq_3$ or a segment restriction thereof, thus $\alpha \leq \gamma$, so the $\leq$ relation on ordinals is transitive and is a linear order. Now let $\mathcal{A}$ be a nonempty set of ordinals. Let $\alpha \in \mathcal{A}$. Let $\leq_1 \in \alpha$ have field A. Consider the set B of all $a \in A$ such that $(\leq_1)_a$ belongs to some element of $\mathcal{A}$. If B is empty, then α is the $\leq$-smallest element of $\mathcal{A}$. If B is nonempty, choose

the smallest a in B: $\mathtt{ot}((\leq_1)_a)$ is the $\leq$-smallest element of $\mathcal{A}$. So the relation $\leq$ on the ordinal numbers is a well-ordering, which is what we set out to prove.

3.11.1 Exercises

1. Some linear orders are listed. For each one, state (correctly) that it is a well-ordering or that it is not. If it is not, explain precisely why it is not (this means give an example of something). If it is, give its order type (an ordinal number).

   (a) $\emptyset$

   (b) the standard order on the integers restricted to $\{x \in \mathbb{Z} \mid -2 \leq x \leq 2\}$

   (c) the standard order on the integers restricted to $\{x \in \mathbb{Z} \mid x \leq 0\}$

   (d) the standard order on the rationals restricted to $\{\frac{n}{n+1} \mid n \in \mathbb{N}\} \cup \{1\}$

   (e) the standard order on the rationals restricted to $\{\frac{n+1}{n} \mid n \in \mathbb{N}\} \cup \{1\}$

   (f) the standard order on the reals restricted to the interval $[0, 1]$

2. Prove that for any natural number n, any two linear orders with a field of size n are isomorphic, and all such linear orders are well-orderings. (How do we prove anything about natural numbers?)

3. Prove that if R and S are well-orderings, so is $R \oplus S$. You need to prove that it is a linear order (which will probably require some reasoning by cases) and prove that it has the additional defining property of a well-ordering.

   Now that you are filled with self-confidence, do the same for $R \otimes S$.

4. Define sets of real numbers such that the restriction of the standard order on the real numbers to that set has each of the following order types:

   (a) $\omega + 1$

   (b) $\omega \cdot 3$

(c) $3 \cdot \omega$

(d) $\omega \cdot \omega$

(e) $\omega \cdot \omega \cdot \omega$ (OK I suppose this is nasty, but see if you can do it)

5. Prove your choice of the two following annoying propositions (these are annoying in the sense that they are straightforward (even "obvious") but there is a good deal to write down).

   (a) Isomorphism is an equivalence relation on relations.

   (b) A relation isomorphic to a well-ordering is a well-ordering.

3.12 Transfinite Induction and Recursion

The following theorem is an analogue of mathematical induction for the ordinals.

Transfinite Induction Theorem: Suppose A is a set of ordinals with the following property: $(\forall \alpha \in \texttt{Ord}.(\forall \beta < \alpha.\beta \in A) \rightarrow \alpha \in A)$. Then $A = \texttt{Ord}$.

Proof: If $A \neq \texttt{Ord}$, then $\texttt{Ord} - A$ is a nonempty set and so contains a least ordinal α. But then obviously $(\forall \beta < \alpha.\beta \in A)$, so $\alpha \in A$ by assumption, which is a contradiction.

Transfinite Induction Theorem (bounded form): Suppose A is a set of ordinals with the following property: $(\forall \alpha < \gamma.(\forall \beta < \alpha.\beta \in A) \rightarrow \alpha \in A)$. Then $(\forall \alpha < \gamma.\alpha \in A)$.

Transfinite Induction Theorem (property form): Suppose $\phi[\alpha]$ is a formula such that $(\forall \alpha \in \texttt{Ord}.(\forall \beta < \alpha.\phi[\beta]) \rightarrow \phi[\alpha])$. Then $(\forall \alpha \in \texttt{Ord}.\phi[\alpha])$.

This looks like the theorem of strong induction for the natural numbers. We can make it look a bit more like the usual formulation of induction by defining some operations on ordinals. The alternative forms are easy to prove and are relevant to untyped set theory where there is no set containing all ordinals. [The property form would have to be restated using a predicate $\texttt{Ord}(x)$ in place of a set of all ordinals to prove theorems about all ordinals in a context where there is no *set* of all ordinals.]

zero: We define 0 as the smallest ordinal (the order type of the empty well-ordering).

successor: For any ordinal α, we define the *successor* of α as the smallest ordinal greater than α. No special notation is needed for successor, since it is easy to show that the successor of α is $\alpha + 1$. Every ordinal has a successor: for any infinite ordinal α containing a well-ordering W with minimal element x, $\mathtt{ot}(W - (\{x\} \times \mathtt{fld}(W)) \cup (\mathtt{fld}(W) \times \{x\}))$ is $\alpha + 1$: the new order is obtained by moving the minimal element of W from bottom to top of the order.

limit ordinal: A nonzero ordinal which is not a successor is called a limit ordinal.

Now we give a different formulation of Transfinite Induction.

Transfinite Induction Theorem: Suppose that A is a set of ordinals such that $0 \in A$, for every ordinal $\alpha \in A$ we also have $\alpha + 1 \in A$, and for any limit ordinal λ such that for all $\beta < \lambda$ we have $\beta \in A$, we also have $\lambda \in A$. Then $A = \mathtt{Ord}$.

Proof: Again, consider the smallest element of the complement of A (there must be a smallest if there is any). It cannot be 0 because $0 \in A$. It cannot be a successor (because its predecessor would be in A, so it would be in A). It cannot be a limit (because everything below it would be in A, so it would be in A). These are the only possibilities.

We now give an extended example of proof by transfinite induction. For purposes of this example, we assume familiarity with the real numbers at the usual undergraduate level. We have seen in an earlier section of these notes how to construct the real numbers in our type theory; mod omitted proofs we are warranted in assuming that they are available at some type and have familiar properties.

Definition: We say that an ordinal α is a *countable ordinal* iff the relations which belong to it have countably infinite fields.

Lemma: For any countable ordinal α, there is a function $f : \mathbb{N} \to \mathtt{Ord}$ such that for natural numbers $i < j$ we have $f(i) < f(j)$, $f(i) < \alpha$ for all i, and α is the least ordinal greater than all $f(i)$'s. More briefly, f is

a strictly increasing sequence of ordinals whose least upper bound is α. We will reserve the right to use the usual notation for sequences, writing $f(i) = \alpha_i$.

Proof of Lemma: Let α be a countable ordinal, and let $\leq$ be a fixed well-ordering of type α with field A. Because α is a countable ordinal, there is an enumeration a_i of the set A (the function $i \in \mathbb{N} \mapsto a_i$ being a bijection from $\mathbb{N}$ to A). We define a sequence b_i recursively as follows: $b_0 = a_0$. Once b_i has been defined as a_j, we define b_{i+1} as a_k, where k is the least natural number such that $a_j < a_k$. The sequence b is strictly increasing, and every element of A is $\leq$-dominated by some element of the range of this sequence (because $a_k \leq b_k$ for every natural number k, as is easy to prove by induction). We can thus define $f(i) = \alpha_i$ as $\mathtt{ot}(\leq)_{b_i}$: these ordinals are clearly all less than the order type α of $\leq$, they increase strictly as the index increases, and any ordinal less than α, being the order type of some $(\leq)_{a_k}$, is dominated by some α_k.

Definition: For any subset X of the interval $(0, 1]$ in the reals and any $a < b$ real numbers, we define $X_{[a,b]}$ as $\{(1 - x)a + xb \mid x \in X\}$. This is a scaled copy of X in the interval $[a, b]$.

For any function f from $\mathbb{N}$ to $\mathcal{P}((0, 1])$ (infinite sequence of subsets of $(0, 1]$), define f^* as $\bigcup\{f(n)_{[1-2^{-n},1-2^{-n-1}]} \mid n \in \mathbb{N}\}$. This construction allows me to put together scaled copies of the infinite sequence of sets $f(n)$, so that the scaled copies are disjoint and appear in the same order that the sets appear in the sequence.

Theorem: For any finite or countable ordinal α, we can find a set of reals $A_\alpha \subseteq (0, 1]$ such that the order type of the restriction of the usual linear order on the reals to A_α is a well-ordering of order type α.

Proof: We break the proof into three cases: $\alpha = 0$, $\alpha = \beta + 1$ for some β, or α a limit ordinal.

In any of these cases, we assume that sets $A_\beta \subseteq (0, 1]$ of reals such that the usual order on the reals restricted to A_β has order type β exist for each ordinal $\beta < \alpha$. Our goal is to show we can find a set of reals A_α such that the order type of the restriction of the usual linear order on the reals to A_α is a well-ordering of order type α.

If $\alpha = 0$, $A_\alpha = \emptyset$ is a subset of the reals such that the restriction of the natural order on the reals to this set has order type $\alpha = 0$.

If $\alpha = \beta + 1$, we assume the existence of A_β as above. The set $(A_\beta)_{[0,\frac{1}{2}]} \cup \{1\}$ has the desired properties: the order type of the natural order on the reals restricted to this set is clearly $\beta + 1$.

If α is a limit ordinal, we have two cases to consider. If α is not a countable ordinal, we have nothing to prove. If α is a countable ordinal, we select a strictly increasing sequence α_i such that the least upper bound of its range is α, as a Lemma above shows we are entitled to do. For each α_i, we are given a set $A_{\alpha_i} \subseteq (0, 1]$ of reals with associated order type α_i. For each i, we select a subset A'_{α_i} of A_{α_i} which we now define. A'_{α_0} is defined as A_{α_0}. For each i, $\leq_{\mathbb{R}} \lceil A_{\alpha_{i+1}}$ has a unique segment restriction of order type α_i. $A'_{\alpha_{i+1}}$ is obtained by subtracting the field of this segment restriction from $A_{\alpha_{i+1}}$. Define $f(i)$ as A'_{α_i} and the set f^* will be the desired set: this is the union of linearly scaled copies of all the A'_{α_i}'s, made successively smaller so they will all fit into $(0, 1]$. It should be clear that the union of such linearly scaled copies has order type α.

Theorem: Any ordinal α which is the order type of the natural order on a subset A of the reals is finite or countable.

Proof: Given such an ordinal α and set A, we construct a set A' such that the natural order on A' also has order type α and all elements of A' are rational numbers (so A' must be finite or countable). For each element $a \in A$, either a is the largest element of A or there is a first element a' of A which is greater than a. This is true because A is well-ordered by the usual order on the reals. Assume that we have an enumeration q_i of the rationals. Let q_a be the first rational in this enumeration which is greater than a and less than a' (or simply the first rational in this enumeration which is greater than a, if a is the largest element of A). It should be evident for all $a, b \in A$ that $q_a < q_b \leftrightarrow a < b$. Thus $\{q_a \mid a \in A\}$ is a set of rationals (thus finite or countable) and the order type of the natural order on this set is α, so α is a finite or countable ordinal.

We conclude that the order types of well-orderings that we can construct as suborders of the natural order on the real numbers are exactly the finite

and countable ordinals. We will see below that there are uncountable ordinals (this will be our first evidence that there are infinite sets which are not countably infinite).

We introduce a type raising operation on ordinals analogous to that already given for cardinals and also traditionally denoted by T.

Definition: For any relation R, we define R^ι as $\{\langle \iota(x), \iota(y)\rangle \mid x\,R\,y\} = \{\langle \{x\}, \{y\}\rangle \mid x\,R\,y\}$. Notice that R^ι is one type higher than R and would seem in some external sense to be isomorphic to R. $R^{\iota^{\mathbf{n}}}$ is similarly defined as $\{\langle \iota^{\mathbf{n}}(x), \iota^{\mathbf{n}}(y)\rangle \mid x\,R\,y\}$

Definition: For any ordinal α, we define $T(\alpha)$ as $\mathtt{ot}(R^\iota)$ for any $R \in \alpha$ (it is easy to show that the choice of R does not matter). Of course we can then also define $T^{\mathbf{n}}(\alpha)$ and $T^{-\mathbf{n}}(\alpha)$ in the natural ways.

Induction can actually be carried out along any well-ordering, but it is traditional to translate all transfinite inductions into terms of ordinals. A general way to do this involves indexing the elements of $\mathtt{fld}(\leq)$ for a general well-ordering $\leq$ with ordinals:

ordinal indexing: For any well-ordering W, define W_α as the unique element x of $\mathtt{fld}(W)$ (if there is one) such that $\mathtt{ot}((\leq)_x) = \alpha$. [Note that if W is a well-ordering of a set of ordinals this is different from $(W)_\alpha$, the segment restriction of W to elements which are W-less than α.]

Notice that the type of α is one higher than the type of W and two higher than the type of W_α (it would be four higher than the type of W_α if we used the Kuratowski pair).

W_α will be defined for each α iff $\alpha < \mathtt{ot}(W)$.

Discussion of ordinal indexing in the natural order on the ordinals themselves requires the following

Theorem: $\mathtt{ot}((\leq_\Omega)_\alpha) = T^{\mathbf{2}}(\alpha)$

Proof: This is proved by transfinite induction. Note that what it says is that the order type of the segment restriction of the natural order on the ordinals to the ordinals less than α is $T^{\mathbf{2}}(\alpha)$. It is "obvious" that this order type is actually α itself, but of course the order type of the segment restriction is two types higher than α itself, so it is seen to be the corresponding ordinal $T^{\mathbf{2}}(\alpha)$ two types higher.

So $[\leq_\Omega]_\alpha = T^{-\mathbf{2}}(\alpha)$ (not α itself).

Note that $[\leq_\Omega]_\alpha$ will be undefined for $\alpha = \mathtt{ot}(\leq_\Omega) = \Omega$, but $[\leq_\Omega]_{T^{\mathbf{2}}(\Omega)} = \Omega$. This shows that $T^{\mathbf{2}}(\Omega)$ is not equal to Ω: in fact $T^{\mathbf{2}}(\Omega) < \Omega$ because $T^{\mathbf{2}}(\Omega)$ is the order type of a segment restriction of the natural order on the ordinals, whose order type is Ω.

The result that $T^{\mathbf{2}}(\Omega) < \Omega$ (in which there is of course a kind of punning reference to the sets of ordinals at different types) shows that there are in effect more ordinals in higher types. There is no well-ordering in type k as long as the natural order on the ordinals in type $k+2$.

Now we prove that there are uncountable ordinals.

Theorem: There are ordinals which are not finite or countably infinite (in sufficiently high types), and so there is in particular a first uncountably infinite ordinal ω_1.

Proof: Consider the restriction of the natural well-ordering on the ordinals to the finite and countable ordinals. This is a well-ordering, so it has an order type, which we call ω_1. For each countable ordinal α, the order type of $(\leq_\Omega)_\alpha$ is $T^{\mathbf{2}}(\alpha)$, and of course $T^{\mathbf{2}}(\alpha) < \omega_1$, because the former is the order type of a segment restriction of the latter. So it cannot be the case that $\omega_1 = T^{\mathbf{2}}(\alpha)$ for any countable ordinal α (of type two lower than that of ω_1). It only remains to show that every countable ordinal of the same type as ω_1 is of the form $T^{\mathbf{2}}(\beta)$. Suppose that γ is a countable ordinal of the same type as ω_1. γ is the order type of some well-ordering $\leq$ with field the set of natural numbers. Now consider $\{\langle T^{-\mathbf{2}}(m), T^{-\mathbf{2}}(n)\rangle \mid m \leq n\}$. We know that there is a set of natural numbers two types lower than the one that $\leq$ orders, because γ is of the same type as ordinals $T^{\mathbf{2}}(\alpha)$ with α countable. We know that the $T^{-\mathbf{1}}$ operation is total on the natural numbers. It follows that the relation just defined makes sense and is of some countable order type β, with $\gamma = T^{\mathbf{2}}(\beta)$, so $\gamma < \omega_1$. But γ is an arbitrary countable ordinal of the type of ω_1, so ω_1 is uncountably infinite.

Corollary: There are sets which are infinite but not countably infinite.

Proof: The field of any relation of type ω_1 will serve: the set of finite and countable ordinals is shown to be uncountably infinite in the argument above.

Here is another very important result about well-orderings whose proof is assisted by ordinal indexing.

Theorem: Suppose that $\leq_1 \subseteq \leq_2$ are well-orderings. Then $\mathtt{ot}(\leq_1) \leq \mathtt{ot}(\leq_2)$.

Proof: We can prove by an easy transfinite induction that $[\leq_2]_\alpha$ is defined and $[\leq_2]_\alpha \leq_2 [\leq_1]_\alpha$ for each ordinal $\alpha < \mathtt{ot}(\leq_1)$. The map taking each $[\leq_1]_\alpha$ to $[\leq_2]_\alpha$ is the desired isomorphism witnessing $\mathtt{ot}(\leq_1) \leq \mathtt{ot}(\leq_2)$.

Of course, when the author says something is easy, that means he or she doesn't really want to take the trouble to prove it. We now do so.

We prove by transfinite induction that $[\leq_2]_\alpha$ is defined and $[\leq_2]_\alpha \leq_2 [\leq_1]_\alpha$ for each ordinal $\alpha < \mathtt{ot}(\leq_1)$.

Note first that an ordinal α is less than $\mathtt{ot}(\leq_1)$ precisely if it is the order type of some $(\leq_1)_x$, by the definition of the order on the ordinals, and this x is $[\leq_1]_\alpha$ by the definition of ordinal indexing, so certainly $[\leq_1]_\alpha$ is defined for every $\alpha < \mathtt{ot}(\leq_1)$.

We fix an ordinal $\alpha < \mathtt{ot}(\leq_1)$. We assume that for every $\beta < \alpha$, $[\leq_2]_\beta$ is defined and $[\leq_2]_\beta \leq_2 [\leq_1]_\beta$. Our goal is to show that $[\leq_2]_\alpha$ is defined and $[\leq_2]_\alpha \leq_2 [\leq_1]_\alpha$.

Observe that $[\leq_1]_\alpha$ exists, and for every $\beta < \alpha$, $[\leq_2]_\beta \leq_2 [\leq_1]_\beta \leq_2 [\leq_1]_\alpha$. ($[\leq_1]_\beta \leq_1 [\leq_1]_\alpha \rightarrow [\leq_1]_\beta \leq_2 [\leq_1]_\alpha$ because $\leq_1 \subseteq \leq_2$). This means that there is at least one object which is $\geq_2$ all the $[\leq_2]_\beta$'s for $\beta < \alpha$, so there must be a $\leq_2$-least such object x. We claim that $x = [\leq_2]_\alpha$. The objects $\leq_2 x$ are precisely the $[\leq_2]_\beta$'s for $\beta < \alpha$, so the order types of the initial segments of $(\leq_2)_x$ are precisely the ordinals less than α, so the ordinals less than the order type of $(\leq_2)_x$ are precisely the ordinals less than α, and so its order type ... is α as desired.

Now we develop a construction analogous to recursive definition of functions of the natural numbers. Just as transfinite induction is analogous to strong induction on the natural numbers, so transfinite recursion is analogous to course-of-values recursion on the natural numbers.

Transfinite Recursion Theorem: We give a nonce definition of $\mathcal{F}$ as the set of all functions whose domains are segments of the natural order on the ordinals [or on the ordinals less than a fixed γ]:

$$\mathcal{F} = \{f \mid (\exists \alpha \in \mathtt{Ord}. f : \mathtt{seg}_{\leq_\Omega}(\alpha) \rightarrow V)\}.$$

Let G be a function from $\mathcal{F}$ to $\iota``V$. Then there is a unique function g with domain $\mathtt{Ord}$ [or with domain the set of ordinals less than γ] with the property that for every ordinal α [or for every ordinal $\alpha < \gamma$], $\{g(\alpha)\} = G(g\lceil\{\beta \mid \beta < \alpha\})$.

Proof: We say that a set I is G-inductive iff whenever a function $f \in \mathcal{F}$ with domain $\{\beta \in \mathtt{Ord} \mid \beta < \alpha\}$ is a subset of I, $\{\alpha\} \times G(f)$ will be a subset of I. Our claim is that g, defined as the intersection of all G-inductive sets, is the desired function.

We first observe that $\mathtt{Ord} \times V$ is G-inductive, so every element of g actually is an ordered pair whose first projection is an ordinal, as we would expect.

We then prove by transfinite induction on α that $g_\alpha = g \cap \mathtt{seg}_{\leq_\Omega}(\alpha) \times V$ is a function with domain $\mathtt{seg}_{\leq_\Omega}(\alpha)$. For $\alpha = 0$ this is obvious (the empty set is a function with domain the empty set of all ordinals less than 0). Suppose that g_β is a function with domain the set of ordinals less than β: our goal is then to show that $g_{\beta+1}$ is a function with domain the set of ordinals less than $\beta + 1$. We claim that $X_\beta = g_\beta \cup (\{\beta\} \times G(g_\beta)) \cup (\{\gamma \mid \gamma > \beta\} \times V)$ is G-inductive. Suppose that f is a function with domain the set of ordinals less than δ and f is a subset of X_β. If $\delta < \beta$, it follows that f is a subset of g_β and so $\{\delta\} \times G(f)$ is a subset of g (because g is G-inductive) and also a subset of g_β and so of X_β because the first projection of its sole element is $\delta < \beta$. If $\delta = \beta$, then $f = g_\beta$ and $\{\beta\} \times G(g_\beta)$ is a subset of X_β by construction. If $\delta > \beta$, then $G(f)$ is a subset of X_β because the first projection of its sole element is $\delta > \beta$. From this we can see that $g_\beta \cup G(g_\beta)$ is precisely $g_{\beta+1}$: G-inductiveness of g shows that $g_\beta \cup (\{\beta\} \times G(g_\beta))$ must be included in g, because g_β is included in g; G-inductiveness of X_β shows that g, and so $g_{\beta+1}$, does not include any ordered pairs with first component $\beta + 1$ and second component outside of $G(g_\beta)$. Clearly $g_{\beta+1}$ is a function, with the same value as g_β at each ordinal $< \beta$ and the sole element of $G(g_\beta)$ as its value at β, so its domain is the set of all ordinals less than $\beta + 1$ as desired. Now we consider the case of a limit ordinal λ with the property that g_β is a function for each $\beta < \lambda$. In this case g_λ is the union of all the g_β's. The only way it could fail to be a function is if some two g_β's had distinct values at some ordinal. But this is impossible: it is clear from the definition that $g_\beta \subseteq g_{\beta'}$ for

$\beta < \beta'$. It is also obvious that the domain of g_λ is the union of the domains of the g_β's, and the union of the segments determined by the ordinals less than a limit ordinal is the segment determined by that limit ordinal.

Since g is a relation with domain the set of ordinals and its restriction to any initial segment of the ordinals is a function, it is a function. We showed above that the value of $g_{\beta+1}$ (g restricted to the ordinals less than $\beta + 1$) at β is the sole element of $G(g_\beta)$, the value of G at the restriction of g to the ordinals less than β, and this is the recurrence relation we needed to show. Suppose that $g \neq g'$ were two distinct functions satisfying this recurrence relation. Let δ be the smallest ordinal such that $g(\delta) \neq g'(\delta)$. Note that $\{g(\delta)\} = G(g\lceil\{\gamma \mid \gamma < \delta\}) = G(g'\lceil\{\gamma \mid \gamma < \delta\}) = \{g'(\delta)\}$ by the shared recurrence relation and the fact that g and g' agree at ordinals less than δ, a contradiction.

We give the qualifications needed for a bounded formulation of recursion in brackets in the statement of the theorem: this is the form which would be used in untyped set theory but also in many applications in typed set theory.

We present a variation of the Recursion Theorem:

Transfinite Recursion Theorem: Suppose we are given a set a, a function f and a singleton-valued function F (of appropriate types which can be deduced from the conclusion): then there is a uniquely determined function $g : \mathtt{Ord} \to V$ such that $g(0) = a$, $g(\alpha + 1) = f(g(\alpha))$ for each α, and $g(\lambda)$ is the sole element of $F(\{g(\beta) \mid \beta < \lambda\})$ for each limit ordinal λ.

Proof: This is a special case of the theorem above. The function $G : \mathcal{F} \to \iota``V$ is defined so that $G(\emptyset) = \{a\}; G(k) = \{f(k(\alpha))\}$ if α is the maximum element of the domain of k; $G(k) = F(\{k(\beta) \mid \beta < \lambda\})$ if the limit ordinal λ is the supremum of the domain of k. The stated recurrence relations are then equivalent to $\{g(\alpha)\} = G(g\lceil\{\beta \mid \beta < \alpha\})$.

The alternative theorem could also be stated in a bounded form.

We define ordinal iteration in a special case. Suppose f is a function and $\leq$ is an order on elements of its field understood from context. Define $f^0(x)$ as x, $f^{\alpha+1}(x)$ as $f(f^\alpha(x))$, and $f^\lambda(x)$ as $\sup\{f^\beta(x) \mid \beta < \lambda\}$. This will uniquely determine a function by either of the recursion theorems. It

would seem most natural to do this construction when f was an increasing function in $\leq$ with the property $x \leq f(x)$. A common choice of $\leq$ would be the subset relation.

The arithmetic operations on the ordinals defined above can also be defined by transfinite recursion.

recursive definition of addition: This resembles the iterative definition of addition on the natural numbers.

1. $\alpha + 0 = \alpha$
2. $\alpha + (\beta + 1) = (\alpha + \beta) + 1$
3. $\alpha + \sup(A) = \sup(\{(\alpha + \beta) \mid \beta \in A\})$

recursive definition of multiplication: This resembles the iterative definition of multiplication on the natural numbers.

1. $\alpha \cdot 0 = 0$
2. $\alpha \cdot (\beta + 1) = \alpha \cdot \beta + \alpha$
3. $\alpha \cdot \sup(A) = \sup(\{\alpha \cdot \beta \mid \beta \in A\})$

recursive definition of exponentiation: Of course a similar definition of exponentiation on natural numbers could be given (and is actually in effect included here). There is a set theoretical definition of exponentiation of ordinals as well, but it is a bit technical.

1. $\alpha^0 = 1$
2. $\alpha^{\beta+1} = \alpha^\beta \cdot \alpha$
3. $\alpha^{\sup(A)} = \sup(\{\alpha^\beta \mid \beta \in A\})$

All the ordinal arithmetic operations commute with the T operation:

Theorem: For any ordinals α and β, $T(\alpha + \beta) = T(\alpha) + T(\beta); T(\alpha \cdot \beta) = T(\alpha) \cdot T(\beta); T(\alpha^\beta) = T(\alpha)^{T(\beta)}$. $T(\alpha) \leq T(\beta) \leftrightarrow \alpha \leq \beta$; if $T^{-\mathbf{1}}(\alpha)$ exists and $T^{-\mathbf{1}}(\beta)$ does not, then $\alpha < \beta$.

We now consider the *original* application of set theory due to Cantor, which includes an example of construction of a function by transfinite recursion. This involves a further discussion of sets of reals.

accumulation point: If X is a set of reals and r is a real number, we say that r is an *accumulation point* of X iff every open interval which contains r contains infinitely many points of X. Note that r does not have to be an element of X to be an accumulation point of X.

closed set: A set of reals X is said to be *closed* iff every accumulation point of X is an element of X.

derived set: For any set X of reals, we define the derived set X' of X as the set of accumulation points of X.

Observations: Obviously X is closed iff $X' \subseteq X$. Whether X is closed or not, X' is closed: if any interval containing r contains infinitely many points of X', then it contains at least one element of X' (accumulation point of X) because it contains infinitely many, and so it contains infinitely many points of X, and so r is itself an accumulation point of X and thus an element of X'. This means further that if we iterate applications of the derived set operator, the first iteration may make our set larger but all subsequent iterations will fix it or remove elements from it.

iteration of the derived set construction: This is a definition by transfinite recursion. Define Δ_0^X as X. Define $\Delta_{\beta+1}^X$ as $(\Delta_\beta^X)'$. At limit stages, take intersections: define Δ_λ^X as $\bigcap\{\Delta_\gamma^X \mid \gamma < \lambda\}$ for each limit ordinal λ.

Theorem: For every countable ordinal α, there is a set of reals $A \subseteq (0,1]$ with the property that $\Delta_\alpha^A = \{1\}$ (and so $\Delta_{\alpha+1}^A = \emptyset$).

Proof: We prove this by transfinite induction on α. If $\alpha = 0$, the set $\{1\}$ has the desired properties. Suppose that we have a set $A \subseteq (0,1]$ such that $\Delta_\beta^A = \{1\}$. Let f be the function which sends each natural number n to the set A: $f^* \cup \{1\}$ will have the desired property. This set consists of infinitely many successively smaller copies of A approaching the limit point $\{1\}$. Application of the derived set operator β times will reduce each of the infinitely many scale copies of A in $f^* \cup \{1\}$ to a single point. The next application of the derived set operator will leave just $\{1\}$ (1 is the only accumulation point). So $f^* \cup \{1\}$ is the desired set for which $\beta + 1$ applications of the derived set operator yields $\{1\}$. Now let λ be a countable limit ordinal. There will be a strictly increasing sequence

λ_i of ordinals such that λ is the least ordinal greater than all the λ_i's (this is proved above). By inductive hypothesis, we may assume that for each i we have a set A_i such that $\Delta^{A_i}_{\lambda_i} = \{1\}$. Define $f(i) = A_i$ (you might note that this actually requires the Axiom of Choice!). Define $A = f^* \cup \{1\}$. Observe that application of the derived set operator to A $\lambda_i + 1$ times eliminates the copy of A_i, for each i. Notice that application of the derived set operator λ_i times always leaves $\{1\}$ in the set, as the scaled copies of A_j for $j > i$ still have nonempty image, so clearly 1 will still be an accumulation point. It follows from these two observations that the intersection of all the sets $\Delta^A_{\lambda_i}$, which will be Δ^A_λ, will contain no element of any of the original scaled copies of the A_i's but will contain 1: it will be $\{1\}$ as required.

The sets shown to exist by this Theorem are in a sense "discrete" (they cannot be dense in any interval, or no iteration of the derived set operation could eliminate them), but have progressively more complex limit structure calibrated by the countable ordinal α. The applications of these concepts by Cantor to problems in the convergence of trignonometric series are the original motivation (or one of the original motivations) for the development of transfinite ordinals and of set theory.

3.12.1 Exercises

1. Prove that for any ordinals α, β, γ if $\alpha + \beta = \alpha + \gamma$ then $\beta = \gamma$."

   You can probably prove this by transfinite induction, using the recursive definitions, but it can be proved using the set theoretic definition and structural properties of ordinals as well.

   Give a counterexample to "if $\beta + \alpha = \gamma + \alpha$ then $\beta = \gamma$.

2. In type theory, prove that for all ordinals α and β, if $\alpha + 1 = \beta + 1$ then $\alpha = \beta$. This is best proved by considering actual well-orderings and isomorphisms between them (not by transfinite induction).

3. Prove by transfinite induction: Every infinite ordinal can be expressed in the form $\lambda + n$, where λ is a limit ordinal and n is a finite ordinal, and moreover it can be expressed in this form in only one way (for this last part you might want to use the result of the previous problem).

3.13 Lateral Functions and T operations; Type-Free Isomorphism Classes

We have observed that cardinals κ and $T^{\mathbf{n}}(\kappa)$, though of different types, are in some sense the same cardinal, and similarly that ordinals α and $T^{\mathbf{n}}(\alpha)$, though of different types, are in some sense the same order type.

We have $T^{\mathbf{n}}(|A|) = |B|$ iff $|\iota^{\mathbf{n}}\text{“}A| = |B|$, that is iff there is a bijection $f : \iota^{\mathbf{n}}\text{“}A \to B$. The bijection f witnesses the fact that A and B are "the same size", by exploiting the fact that A and $\iota^{\mathbf{n}}\text{“}A$ are externally "the same size".

We introduce the following definitions.

Definition (lateral relations): If $R \subseteq \iota^{\mathbf{n}}\text{“}A \times B$, we define $x\, R_{\mathbf{n}}\, y$ as holding iff $\iota^{\mathbf{n}}(x)\, R\, y$. Similarly, if $S \subseteq A \times \iota^{\mathbf{n}}\text{“}B$, we define $x\, S_{-\mathbf{n}}\, y$ as holding iff $x\, S\, \iota^{\mathbf{n}}(y)$.

Definition (description of lateral relations): We define $A \times_{\mathbf{n}} B$ as $\iota^{\mathbf{n}}\text{“}A \times B$ and
$A \times_{-\mathbf{n}} B$ as $A \times \iota^{\mathbf{n}}\text{“}B$.

Definition (lateral functions): If $f : \iota^{\mathbf{n}}\text{“}A \to B$, we define $f_{\mathbf{n}}(a) = f(\iota^{\mathbf{n}}(a))$ for each $a \in A$. Similarly, if $g : A \to \iota^{\mathbf{n}}\text{“}B$, we define $g_{-\mathbf{n}}(a) = \iota^{-\mathbf{n}}(g(a))$.

Definition (description of lateral functions): $f_{\mathbf{n}} : A \to B$ is defined as $f : \iota^{\mathbf{n}}\text{“}A \to B$; $f_{-\mathbf{n}} : A \to B$ is defined as $f : A \to \iota^{\mathbf{n}}\text{“}B$.

Note that in none of these notations is a boldface subscript actually part of the name of a function or relation: the boldface subscripts are always indications of the role the function or relation is playing in the expression.

This definition allows us to code relations and functions with domains and ranges of different types. Note that this definition allows us to say that $T^{\mathbf{n}}(|A|) = |B|$ iff there actually is a (lateral) bijection from A to B! The definition also allows us to assert that well-orderings of types α and $T^{\mathbf{n}}(\alpha)$ actually are "isomorphic" in the sense that there is a lateral function satisfying the formal conditions to be an isomorphism between them.

We present the Transfinite Recursion Theorem in a slightly different format:

Transfinite Recursion Theorem: We give a nonce definition of $\mathcal{F}$ as the set of all functions whose domains are segments of the natural order on the ordinals [or on the ordinals less than a fixed γ]. Let $G_{-\mathbf{1}} : \mathcal{F} \to \mathcal{V}$. Then there is a unique function g with domain $\mathtt{Ord}$ [or with domain the set of ordinals less than γ] with the property that for every ordinal α [or for every ordinal $\alpha < \gamma$], $g(\alpha) = G_{-\mathbf{1}}(g\lceil \mathtt{seg}(\alpha))$.

We give a general "comprehension" theorem for functions and relations with a type differential.

Theorem: If $\phi[x^n, y^{n+k}]$ is a formula, there is a set relation R such that $x\, R_{\mathbf{k}}\, y \leftrightarrow \phi[x, y]$ (where types revert to being implicit in the second formula).

If $\phi[x^{n+k}, y^n]$ is a formula, there is a set relation R such that $x\, R_{-\mathbf{k}}\, y \leftrightarrow \phi[x, y]$ (where types revert to being implicit in the second formula).

If $(\forall x^n \in A.(\exists! y^{n+k}.\phi[x^n, y^{n+k}]))$, then there is a function $f_{\mathbf{k}} : A \to V$ such that for any $x \in A$, $y = f_{\mathbf{k}}(x) \leftrightarrow \phi[x, y]$.

If $(\forall x^{n+k} \in A.(\exists! y^n.\phi[x^{n+k}, y^n]))$, then there is a function $f_{-\mathbf{k}} : A \to V$ such that for any $x \in A$, $y = f_{-\mathbf{k}}(x) \leftrightarrow \phi[x, y]$.

Corollary: If A^n and B^{n+k} are sets and there is a formula $\phi[a, b]$ such that $(\forall a \in A.(\exists! b \in B.\phi[a, b])) \wedge (\forall b \in B.(\exists! a \in A, \phi[a, b]))$, then $T^k(|A|) = |B|$. If $\leq_1^n$ and $\leq_2^{n+k}$ are well-orderings, and there is a formula ϕ such that $(\forall xy.x <_1 y \leftrightarrow (\exists zw.\phi[x, z] \wedge \phi[y, w] \wedge z <_2 w)) \wedge (\forall zw.z <_2 w \leftrightarrow (\exists xy.\phi[z, x] \wedge \phi[w, y] \wedge x <_1 y))$, then $T^k(\mathtt{ot}(\leq_1)) = \mathtt{ot}(\leq_2)$.

All parts of this theorem are proved by direct application of the Axiom of Comprehension. The Corollary expresses the idea that any external bijection or isomorphism we can describe using a formula is actually codable by a set and so witnesses appropriate cardinal or ordinal equivalences.

We note that T operations can be defined for general isomorphism classes.

Definition: For any relation R, the isomorphism class $[R]_\approx = \{S \mid R \approx S\}$. We define $T([R]_\approx) = [R^\iota]_\approx$, where $R^\iota = \{\langle \{x\}, \{y\}\rangle \mid x\, R\, y\}$, as already defined. Note that this is more general than but essentially the same as the T operation on ordinals.

Now we pursue an extension of the Reasonable Convention proposed above for natural numbers. We recall that the T operation on cardinals witnesses an exact correspondence between the natural numbers at different types. This allows us, if we wish, to introduce natural number variables which can be used in a type-free manner: such a variable can be shifted into the type appropriate for any context by appropriate applications of the T operation or its inverse. All statements purely in the language of the natural numbers are invariant under uniform application of the T operation, as we have seen. Each occurrence of a natural number variable translates into an occurrence of a general variable of an appropriate type restricted to the set of natural numbers at the appropriate type.

This idea can be extended to cardinals and ordinals (and to isomorphism classes in general), but a further refinement is needed. The difficulty is that the ordinals in one type are mapped injectively into but not onto the ordinals in the next type, as we have just seen. We will see below that the same is true of the cardinals. The natural number variables introduced in the previous paragraph are translated as general variables restricted to the set of all natural numbers (which is in effect the same set at each type); this cannot work for the ordinals (or the cardinals): each ordinal bound variable must be restricted to the ordinals in a specific type (which is equivalent to restriction to an initial segment of "all the ordinals" determined by the first ordinal not in that particular type (the first ordinal of the next higher type which is not an image under T)). We can thus use type-free ordinal variables as long as we require that any such variable be restricted to a proper initial segment of the ordinals (the type of the bound will determine the highest type in which we can be working), and we can treat cardinals similarly. There is no way to express a general assertion about all ordinals at whatever type in type theory. Just as in natural number arithmetic, all statements about properties, relations, and operations natural to cardinals and ordinals are invariant under uniform application of the T operation: this enables the proposed identifications of cardinals and ordinals at diverse types to cohere.

This convention would allow the elimination in practice of the inconvenient reduplication of cardinals, ordinals, and similar constructions at each type. We do not use it as yet, but it is important to us to note that it is possible to use this convention.

3.14 Other Forms of the Axiom of Choice

The Axiom of Choice is equivalent to some other interesting propositions (in fact, there are whole books of them but we will only discuss a few).

The Well-Ordering Theorem: Every set is the field of a well-ordering. (Equivalently, V is the field of a well-ordering.)

Observation: It is obvious that the well-ordering theorem implies the Axiom of Choice: the choice set of a partition can be taken to be the set of minimal elements in the elements of the partition under a well-ordering of the union of the partition. The interesting part of the result is the converse: the Axiom of Choice is enough to prove the Well-Ordering Theorem.

Definition: A *chain in a partial order* $\leq$ is a subset C of $\mathtt{fld}(\leq)$ such that $\leq \cap C^2$, the restriction of $\leq$ to C, is a linear order (i.e., any two elements of C are comparable in the order).

Definition: A collection of sets is said to be *nested* iff it is a chain in the inclusion order: A is a nested collection of sets iff $(\forall x \in A.(\forall y \in A.x \subseteq y \vee y \subseteq x))$.

Lemma: The union of a nested collection of chains in a partial order $\leq$ is a chain in $\leq$.

Zorn's Lemma: A partial order with nonempty domain in which every chain has an upper bound has a maximal element.

Observation: Let $\mathcal{A}$ be the set of all well-orderings of subsets of a set A. We define $U \leq V$ as holding for $U, V \in \mathcal{A}$ iff either $U = V$ or U is a segment restriction of V. A chain in this well-ordering is a collection C of well-orderings of A which agree with one another in a strong sense and whose union will also be a well-ordering of a subset of A and so an upper bound of the chain C (details of this bit are left as an exercise). So Zorn's Lemma would allow us to conclude that there was a maximal partial well-ordering of A under the segment restriction relation, which clearly must be a well-ordering of all of A (any element not in the field of the maximal well-order could be adjoined as a new largest element of a larger well-ordering for a contradiction).

Since Zorn implies Well-Ordering and Well-Ordering implies Choice, it only remains to show that Choice implies Zorn to prove that all three are equivalent (in the presence of the rest of our axioms).

Proof of Zorn's Lemma: Let $\leq$ be a partial order in which every chain has an upper bound.

Let $\mathcal{C}$ be the set of all chains in $\leq$. Note that for any chain C if there is an upper bound of C which belongs to C there is exactly one such upper bound. If in addition all upper bounds of C belong to C then this uniquely determined upper bound is maximal in $\leq$. For each chain C in $\leq$, define B_C as the set of all upper bounds for C which are not in C, if there are any, and otherwise as the singleton of the unique upper bound of C which is an element of C. All of these sets will be nonempty if $\leq$ has no maximal element. The set $\{\{C\} \times \iota\text{“}B_C \mid C \in \mathcal{C}\}$ is a partition, and so has a choice set. Notice that the choice set is a function F which sends each $C \in \mathcal{C}$ to the singleton set of an upper bound of C, which will belong to C only if all upper bounds of C belong to C (in which case the upper bound is maximal).

For each chain C, denote the linear order $\leq \cap C^2$ by $\leq_C$. We call a chain C a *special chain* iff $\leq_C$ is a well-ordering and for each $x \in C$ we have $\{x\} = F(\mathtt{fld}((\leq_C)_x))$.

We can prove by transfinite induction that $\leq_C$ is precisely determined by its order type (for any special chains C and D, if $\leq_C$ is isomorphic to $\leq_D$ then $\leq_C = \leq_D$). Suppose otherwise: then there is a least ordinal to which distinct $\leq_C$ and $\leq_D$ belong. There must be a $\leq_C$-first element x which differs from the corresponding $\leq_D$ element y. But this implies that $(\leq_C)_x = (\leq_D)_y$ whence $\{x\} = F((\leq_C)_x) = F((\leq_D)_y) = \{y\}$.

This implies further that for any two distinct special chains, one is a segment restriction of the other. This further implies that the union of all special chains is a linear order and in fact a special chain; call it E. Now $E \cup F(\leq_E)$ is a special chain as well, which cannot properly extend E, so $F(\leq_E) \subseteq E$, so the sole element of $F(E)$ is a maximal element with respect to $\leq$.

Alternative Proof of Zorn's Lemma: Let $\leq$ be a nonempty partial order in which any chain has an upper bound. Let $\mathcal{C}$ be the set of all chains in $\leq$.

For each chain C in $\leq$ and $x \in \mathtt{fld}(\leq)$, we say that x is an appropriate upper bound of C if x is an upper bound of C and $x \notin C$ or if $x \in C$ and all upper bounds of C are elements of C. Notice that if there is an upper bound of C belonging to C there is only one, and also notice that if the unique upper bound of C belonging to C is the only upper bound of C then it is maximal in $\leq$, because anything strictly greater than the unique upper bound of C in C would be an upper bound of C not in C.

For each chain C in $\leq$, we define X_C as the set of all ordered pairs $\langle C, \{x\}\rangle$ such that x is an appropriate upper bound of C. Notice that if $C \neq C'$ then X_C and $X_{C'}$ are disjoint (because elements of the two sets are ordered pairs with distinct first projections). Thus $\{X_C \mid C \in \mathcal{C}\}$ is a partition, and has a choice set F. Notice that F is a function, $F : \mathcal{C} \to \iota\text{“}\mathtt{fld}(\leq)$, and $F(C)$ for every C is the singleton $\{x\}$ of an appropriate upper bound x of C.

Define a function G by transfinite recursion: $G(\alpha)$ is defined as the sole element of $F(\mathtt{rng}(G\lceil\{\beta \mid \beta < \alpha\}))$ if $\mathtt{rng}(G\lceil\{\beta \mid \beta < \alpha\})$ is a chain in $\leq$ and as 0 otherwise. Transfinite induction shows that $\mathtt{rng}(G\lceil\{\beta \mid \beta < \alpha\})$ is a chain in $\leq$ for any ordinal α (for successor α, because $C \cup F(C)$ is always a chain in $\leq$ if C is a chain in $\leq$, and for limit α because a union of nested chains in $\leq$ is a chain in $\leq$). Note that $G(\alpha)$ will not be one of the $G(\beta)$'s for $\beta < \alpha$ unless it is maximal for $\leq$, so G is injective if $\leq$ has no maximal element. The range of G is a subset of the field of $\leq$ with a unique well-ordering under which G is an increasing function. The order type of this well-ordering will be the order type Ω of the ordinals iff G is injective. If G is not injective, it is constant past a certain point and so the order type of this well-ordering will be that of an initial segment of the ordinals, so strictly less than Ω. Now we employ a trick: consider instead of $\leq$ the order type $\leq^{\iota^{\mathbf{2}}}$ of double singletons induced by $\leq$. The well-ordering of the range of the function G associated with $\leq^{\iota^{\mathbf{2}}}$ will have some order type $T^{\mathbf{2}}(\alpha) < \Omega$ (because it is a well-ordering of a set of double singletons) and so cannot be injective, and so $\leq^{\iota^{\mathbf{2}}}$ has a maximal element, from which it follows that $\leq$ itself has a maximal element. The point of the trick is that the original working type we started with might not have had enough ordinals for the construction of G to exhaust the field of $\leq$.

Throughout this discussion we could have used the lateral function notation introduced in the previous subsection: $F_{-\mathbf{1}}(C)$ is an upper bound for C for each chain C.

NOTE: include examples of use of Zorn's Lemma in other parts of mathematics.

The Axiom of Choice directly enables us to make choices from pairwise disjoint collections of sets. But in fact we can use the Axiom to show that we can make choices from any collection of nonempty sets.

Definition: Let A be a collection of nonempty sets. A function c with domain A is called a *choice function for* A iff $c : A \to 1$ ($c(a)$ is a one element set for each $a \in A$) and $c(a) \subseteq a$ for each $a \in A$. The sole element of $c(a)$ is the item selected from A by the choice function.

It is equivalent to say (using the notation for lateral functions) that a choice function for A is a function $c_{-\mathbf{1}} : A \to V$ such that $c_{-\mathbf{1}}(a) \in a$ for each $a \in A$.

Theorem: Each collection of nonempty sets A has a choice function.

Proof: The collection $\{\{a\} \times \iota\text{“}a \mid a \in A\}$ is a partition and so has a choice set c. This choice set is the desired choice function.

We define a logical device which will prove useful later.

Hilbert symbol: Let H be a fixed function $V \to 1$ such that $H\lceil(V - \{\emptyset\})$ is a choice function. We do not care which one. Define $(\epsilon x.\phi)$ as the sole element of $H(\{x \mid \phi\})$ for each formula ϕ.

Theorem: For any formula ϕ, $(\exists x.\phi) \leftrightarrow \phi[(\epsilon x.\phi)/x]$. Since $(\forall x.\phi) \leftrightarrow \neg(\exists x.\neg\phi)$, this means that both quantifiers could be defined in terms of the Hilbert symbol.

Proof: This is obvious.

Note that a systematic use of the Hilbert symbol would imply a choice of an H in each relevant type.

3.14.1 Exercises

1. Prove that the union of a nested set of chains in a partial order $\leq$ is a chain. A chain is a set C such that for any $x, y \in C$ we have either $x \leq y$ or $y \leq x$; a nested collection of sets is a set A of sets which is a chain in the subset relation (for any $x, y \in A$, either $x \subseteq y$ or $y \subseteq x$).

2. Prove that the union of a countably infinite collection of countably infinite sets is countably infinite. Notice that you already know that $\mathbb{N} \times \mathbb{N}$ is a countable set.

   We give the result in more detail: suppose that F is a function with domain $\mathbb{N}$ and the property that each $F(n)$ is a countably infinite set. Show that $\bigcup\{F(n) \mid n \in \mathbb{N}\}$ is countable (that is, show that it is the range of a bijection with domain the set of natural numbers).

   Hint: be very careful. It is fairly easy to see why this is true if you understand why $\mathbb{N} \times \mathbb{N}$ is a ocuntable set, but there is an application of the Axiom of Choice involved which you need to notice; in type theory or set theory without choice there may be countable collections of countable sets which have uncountable unions!

3. Use Zorn's Lemma to prove that every infinite set is the union of a pairwise disjoint collection of countably infinite sets.

   Then prove that if B is a collection of countably infinite sets, $|\bigcup B| = |\bigcup B| + |\bigcup B|$. (This exploits the fact that $|\mathbb{N}| = |\mathbb{N}| + |\mathbb{N}|$; it also requires the Axiom of Choice).

   Notice that this is another proof that $\kappa + \kappa = \kappa$ for any infinite cardinal κ.

3.15 Transfinite Arithmetic of Order, Addition, and Multiplication

We define the order relation on cardinals in a natural way.

order on cardinals: $|A| \leq |B|$ iff there is an injection from A to B.

Implicit in our notation is the claim that $\leq$ is a partial order. The relation is obviously reflexive and transitive: that it is antisymmetric is a famous theorem.

Cantor-Schröder-Bernstein Theorem: If $|A| \leq |B|$ and $|B| \leq |A|$ then $|A| = |B|$.

Before proving this theorem we give an example to illustrate why it is not obvious. Consider the sets $[0,1] = \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$ and $\mathcal{P}(\mathbb{N})$, the set of all sets of natural numbers.

A injection f from $[0,1]$ into $\mathcal{P}(\mathbb{N})$ is defined by $f(r) = \{k \in \mathbb{N} \mid \frac{1}{2^{k+1}}$ is a term in the unique nonterminating binary expansion of $r\}$ while a bijection g from $\mathcal{P}(\mathbb{N})$ into $[0,1]$ is given by $g(A) = \Sigma_{k \in A} \frac{1}{10^{k+1}}$. So it is easy to see that each set embeds injectively in the other, but it is not at all easy to see how to construct a bijection which takes one set exactly onto the other.

We now give the slightly delayed

Proof of the Cantor-Schröder-Bernstein Theorem: Assume that there is an injection $f : A \to B$ and an injection $g : B \to A$: our goal is to show that there is a bijection h from A to B. B is the same size as $g\text{“}B \subseteq A$, so if we can show $A \sim g\text{“}B$ we are done. The map $f|g$ sends all of A into $g\text{“}B$; we develop a trick to send it exactly onto $g\text{“}B$. Let C be the intersection of all sets which contain $A - g\text{“}B$ and are closed under $f|g$. Let h_0 be the map which sends all elements of C to their images under $f|g$ and fixes all elements of $A - C$. This is a bijection from A to $g\text{“}B$, so $h_0|g^{-1}$ is a bijection from A to B.

Note that this proof does not use the Axiom of Choice. Beyond this point we will use the Axiom of Choice freely, and some of the results we state are not necessarily true in type theory or set theory without Choice.

Theorem: The natural order on cardinals is a linear order.

Proof: Let A and B be sets: we want to show $|A| \leq |B|$ or $|B| \leq |A|$. This is easy using the Well-Ordering Theorem: we choose well-orderings $\leq_A$ and $\leq_B$ of A and B respectively. If the well-orderings are isomorphic, the isomorphism between them witnesses $|A| = |B|$ (and so $|A| \leq |B|$. Otherwise, one of $\leq_A$ and $\leq_B$ is isomorphic to a segment restriction of the other, and the isomorphism is the required injection from one of the sets into the other.

Theorem: The natural order on cardinals is a well-ordering.

Proof: Let C be a set of cardinals. Our aim is to show that C has a smallest element in the natural order. Let $\leq$ be a well-ordering of a set at least as large as any of the elements of the union of C (the universe of the appropriate type will work). Consider the set of all well-orderings of elements of the union of C (note that the union of C is the set of all sets which have cardinalities in the set C). Every well-ordering in this set will either be similar to $\leq$ or similar to some segment restriction of $\leq$. If all are similar to $\leq$, then all elements of C are the same and it has a smallest element. Otherwise consider the set of all x such that $\leq_x$ is isomorphic to some well-ordering of some element of the union of C: there must be a $\leq$-smallest element of this set, which corresponds to the smallest element of C in the natural order.

Theorem: There is a surjection from A onto B iff $|B| \leq |A|$ (and B is nonempty if A is).

Proof: If there is an injection f from B to A, then we can define a surjection from A to B as follows: choose $b \in B$; map each element of A to $f^{-1}(a)$ if this exists and to b otherwise. This will be a surjection. If B is empty we cannot choose b, but in this case A is empty and there is obviously a surjection.

If there is a surjection f from A onto B, there is a partition of A consisting of all the sets f"$\{a\}$ for $a \in A$. Let C be a choice set for this partition. Map each element b of B to the unique element of $C \subseteq A$ which is sent to b by f. This map is obviously an injection.

Definition: In type theory or set theory *without* Choice, we define

$$|A| \leq^* |B|$$

as holding iff there is a bijection from B onto A. In the light of the previous Theorem, there is no need for this notation if we assume Choice.

Theorem: For all cardinals κ and λ, $\kappa \leq \lambda \leftrightarrow T(\kappa) \leq T(\lambda)$. If $T^{-\mathbf{1}}(\kappa)$ exists and $T^{-\mathbf{1}}(\lambda)$ does not exist then $\kappa \leq \lambda$.

Definition (repeated from above): We define $\aleph_0$ as $|\mathbb{N}|$. Elements of $\aleph_0$ are called *countably infinite sets*, or simply *countable sets*.

Theorem: $\aleph_0 + 1 = \aleph_0 + \aleph_0 = \aleph_0 \cdot \aleph_0 = \aleph_0$. It is straightforward to define a bijection between $\mathbb{N}$ and $\mathbb{N} \times \mathbb{N}$. The bijections between the naturals and the even and odd numbers witness the second statement. The successor map witnesses the first statement.

Theorem: $\aleph_0 = T(\aleph_0)$.

Proof: This follows from the fact that natural numbers are sent to natural numbers by the T operation.

Theorem: Every infinite set has a countable subset.

Proof: Let A be an infinite set. The inclusion order on the collection of all bijections from initial segments of $\mathbb{N}$ to A satisfies the conditions of Zorn's Lemma and so has a maximal element. If the maximal element had domain a proper initial segment of $\mathbb{N}$, then the set would be finite. So the maximal element is a bijection from $\mathbb{N}$ to a subset of A.

Theorem: For every infinite cardinal κ, $\kappa + 1 = \kappa$.

Proof: Let A be an infinite set. The inclusion order on the set of all bijections from B to $B \cup \{x\}$, where $B \cup \{x\} \subseteq A$ and $x \notin B$, satisfies the conditions of Zorn's Lemma and so has a maximal element. It is nonempty because A has a countable subset. If the maximal element is a map from B to $B \cup \{x\}$ and there is $y \in A - (B - \{x\})$, then affixing $\langle y, y \rangle$ to the map shows that the supposed minimal element was not minimal.

Corollary: If n is finite and κ is an infinite cardinal then $\kappa + n = \kappa$.

Theorem: For every infinite cardinal κ, $\kappa + \kappa = \kappa$.

Proof: Let A be an infinite set. The inclusion order on pairs of bijections f and g with domain $B \subseteq A$ and disjoint ranges whose union is B, ordered by componentwise inclusion, satisfies the conclusion of Zorn's Lemma. It is nonempty because A has a countable subset. Suppose that a maximal such pair of bijections has been constructed. If there is no countably infinite subset in $A - B$, then $A - B$ is finite and $|B| = |A|$ by the previous result (iterated) and the result is proved: otherwise take a countable subset of $A - B$ and extend the supposedly maximal map to a larger one.

Corollary: If $\lambda \leq \kappa$ and κ is an infinite cardinal then $\kappa + \lambda = \kappa$.

Theorem: For every infinite cardinal κ, $\kappa \cdot \kappa = \kappa$.

Proof: Let A be an infinite set. The inclusion order on bijections from $B \times B$ to B, where $B \subseteq A$, satisfies the conditions of Zorn's Lemma. It is nonempty because A has a countable subset. Now consider a maximal function in this order, mapping $B \times B$ to B. If $A - B$ contains no subset as large as B, then $|B| = |A|$ by the previous result and the result is proved. Otherwise, choose $B' \subseteq A - B$ with $|B'| = |B|$. It is then easy to see from assumptions about B and B' and the previous result that the map from $B \times B$ to B can be extended to a bijection from $(B \cup B') \times (B \cup B')$ to $B \cup B'$, contradicting the supposed maximality of the bijection.

Corollary: If $\lambda \leq \kappa$ and κ is an infinite cardinal then $\kappa \cdot \lambda = \kappa$.

The arithmetic of addition and multiplication of infinite cardinals is remarkably simple. This simplicity depends strongly on the presence of Choice.

3.15.1 Exercises

1. A classical argument that $|\mathcal{R}^2| = |\mathcal{R}|$ goes as follows. Suppose that it is granted that $|[0,1]| = |\mathcal{R}|$ (this takes a wee bit of work, too, but not too much). So it suffices to prove that $|[0,1]|^2 = |[0,1]|$. Map the pair of numbers with decimal expansions $0.a_1a_2a_3\ldots$ and $0.b_1b_2b_3\ldots$ to the number with expansion $0.a_1b_1a_2b_2a_3b_3\ldots$. Unfortunately, this doesn't quite give us the bijection we want due to problems with decimal expansions (explain). Give a corrected description of this map, taking into account bad features of decimal expansions, and explain why it is not a bijection from $[0,1]^2$ to $[0,1]$. Is it an injection? A surjection? Then use a theorem from the notes (giving all details of its application to this situation) to conclude that there is a bijection from $[0,1]^2$ to $[0,1]$.

3.16 Cantor's Theorem

3.16.1 Cardinal Exponentiation and the Theorem

In this section, we start by defining another arithmetic operation. We have delayed this because its properties in the transfinite context are more vexed.

Definition (function space): The set of all functions from A to B is called B^A. Note that B^A is one type higher than A or B (it would be three types higher if we were using the Kuratowski pair).

Definition (cardinal exponentiation): We define $|A|^{|B|}$ as $T^{-\mathbf{1}}(A^B)$.

The appearance of $T^{-\mathbf{1}}$ is required to get a type-level operation (it would be $T^{-\mathbf{3}}$ if we used the Kuratowski pair). It makes it formally possible that this operation is partial – and indeed it turns out that this operation *is* partial.

Definition: For each subset B of A define χ^A_B as the function from A to $\{0,1\}$ which sends each element of B to 1 and each element of $A - B$ to 0. We call this *the characteristic function of B (relative to A).*

Observation: The function sending each $B \subseteq A$ to χ^A_B is a bijection.

Theorem: $|\mathcal{P}(A)| = |\{0,1\}^A|$, so $2^{|A|} = T^{-\mathbf{1}}(|\mathcal{P}(A)|)$.

Now comes the exciting part.

Cantor's Theorem: For any set A, there is no function f from $\iota``A$ onto $\mathcal{P}(A)$.

Proof: Suppose otherwise, that is, that there is a function f from $\iota``A$ onto $\mathcal{P}(A)$. Consider the set

$$R = \{a \in A \mid a \notin f(\{a\})\}.$$

Since f is onto, $R = f(\{r\})$ for some $r \in A$. Now

$$r \in R \leftrightarrow r \notin f(\{r\}) = R$$

is a contradiction.

This tells us that a set A cannot be the same size as its power set. The fact that A and $\mathcal{P}(A)$ are of different types necessitates the exact form of the theorem. This implies that if $2^{|A|}$ exists, that

$$T(|A|) = |\iota\text{“}A| \neq |\mathcal{P}(A)| = T(2^{|A|})$$

so $|A| \neq 2^{|A|}$. There are at least two distinct infinite cardinals, $|\mathbb{N}|$ and $2^{|\mathbb{N}|}$ (in high enough types for both to be present).

Since certainly $|\iota\text{“}A| \leq |\mathcal{P}(A)|$ (singletons of elements of A are subsets of A), it follows by Cantor-Schroder-Bernstein that $|\mathcal{P}(A)| \not\leq |\iota\text{“}A|$, as otherwise these two cardinals would be equal, so we can write $|\iota\text{“}A| < |\mathcal{P}(A)|$ and $\kappa < 2^\kappa$.

Further we have the curious result that $|\iota\text{“}V| < |\mathcal{P}(V)|$ must be distinct: there are more sets in any given type than singletons of sets (of the next lower type). This implies that V in any given type is strictly larger than any set of lower type (in the sense that the elementwise image under an appropriate $\iota^{\mathbf{n}}n$ of the lower type set at the same type as V will have smaller cardinality than V): $T^{-\mathbf{1}}(|V|)$ is undefined and so is $2^{|V|}$, which would be $T^{-\mathbf{1}}(|\mathcal{P}(V)|)$.

3.16.2 Applications to the Number Systems

We give some set theoretical facts about familiar number systems.

Theorem: $\mathbb{Z} \sim \mathbb{N}$

Proof: Consider the map which sends 0 to 0, $2n-1$ to n for each natural number $n > 0$ and $2n$ to $-n$ for each $n > 0$. This is a bijection.

Theorem: $\mathbb{Q} \sim \mathbb{N}$.

Proof: There is an obvious injection from $\mathbb{Q}$ into $\mathbb{Z} \times \mathbb{N}^+$ determined by simplest forms of fractions. $\mathbb{Z} \times \mathbb{N}^+ \sim \mathbb{N} \times \mathbb{N}$ is obvious. $\mathbb{N} \times \mathbb{N}$ is injected into $\mathbb{N}$ by the map $f(m,n) = 2^m 3^n$, and of course $\mathbb{N}$ injects into $\mathbb{Q}$. The result follows by the Cantor-Schröder-Bernstein theorem.

Theorem: $\mathbb{R} \sim \mathcal{P}(\mathbb{N})$, so $|\mathbb{R}| > |\mathbb{N}|$.

Proof: An injection from the interval $[0,1)$ in the reals into $\mathcal{P}(\mathbb{N})$ is defined by sending each real r in that interval to the set of all natural numbers i such that there is a 1 in the $\frac{1}{2^i}$'s place in the binary expansion of r

which contains only finitely many 1's. An injection from $\mathcal{P}(\mathbb{N})$ to the interval $[0,1)$ sends each set A of natural numbers to the real number whose base 3 expansion consists of 1's in the $\frac{1}{3^i}$'s place for each $i \in A$ and zeroes in all other places. It follows by Cantor-Schröder-Bernstein that

$$[0,1) \sim \mathcal{P}(\mathbb{N}).$$

Injections from $(-\frac{\pi}{2}, \frac{\pi}{2})$ into $[0,1)$ and vice versa are easy to define, so

$$(-\frac{\pi}{2}, \frac{\pi}{2}) \sim [0,1).$$

Finally, the arc tangent function witnesses

$$(-\frac{\pi}{2}, \frac{\pi}{2}) \sim \mathbb{R},$$

The cardinal inequality follows from Cantor's Theorem.

The linear orders on $\mathbb{Q}$ and $\mathbb{R}$ share a characteristic which might suggest to the unwary that both sets should be larger than the "discrete" $\mathbb{N}$.

Definition: If $\leq$ is a linear order and $A \subseteq \texttt{fld}(\leq)$, we say that A is *dense in* $[\leq]$ iff for each $x < y$ there is $z \in A$ such that $x < z$ and $z < y$ (it is traditional to write $x < z \wedge z < y$ as $x < z < y$). We say that $\leq$ itself is merely *dense* iff $\texttt{fld}(\leq)$ is dense in $[\leq]$.

Observation: The natural orders on $\mathbb{Q}$ and $\mathbb{R}$ are dense. $\mathbb{Q}$ is dense in the order on $\mathbb{R}$.

Definition: A linear order with a finite or countable dense set is said to be *separable*. The immediately preceding example shows that a separable linear order need not be countable.

Theorem: Any two dense linear orders with countably infinite field and no maximum or minimum element are isomorphic. This is a characterization of the order on $\mathbb{Q}$ up to isomorphism.

Proof: Let $\leq_1$ and $\leq_2$ be two such orders. Let $\leq^1$ and $\leq^2$ be well-orderings of order type ω with the same fields as $\leq_1$ and $\leq_2$ respectively.

We define a map f from $\texttt{fld}(\leq_1)$ to $\texttt{fld}(\leq_2)$ by a recursive process.

Initially, we define a_1^0 as the $\leq^1$-least element of $\mathtt{fld}(\leq_1)$, and define $f(a_1^0)$ as the $\leq^2$-least element of $\mathtt{fld}(\leq_2)$. This completes stage 0 of the construction.

Suppose that the values at which f has been defined at the nth stage of our construction are the terms $a_i^n (0 \leq i \leq N)$ of a finite strictly $\leq_1$-increasing sequence of elements of $\mathtt{fld}(\leq_1)$, and further that $f(a_i^n)(0 \leq i \leq N)$ is a strictly increasing $\leq_2$- sequence. We define a_{2i+1}^{n+1} as a_i^n for each i in the domain of a^n. Note that this means that f is already defined at each of the odd-indexed elements of the range of a^{n+1} that we will consider in what follows. We define a_0^{n+1} as the $\leq^1$-least element of the $\leq_1$-interval $(-\infty, a_1^{n+1})$ and $f(a_0^{n+1})$ as the $\leq^2$-least element of the $\leq_2$-interval $(-\infty, f(a_1^{n+1}))$. We define a_{2N+2}^{n+1} as the $\leq^1$-least element of the $\leq_1$-interval (a_{2N+1}^{n+1}, ∞), and $f(a_{2N+2}^{n+1})$ as the $\leq^2$-least element of the $\leq_2$-interval $(f(a_{2N+1}^{n+1}), \infty)$. These selections succeed because neither order has a maximum or minimum. For $0 \leq i < N$, we define a_{2i}^{n+1} as the $\leq^1$-least element of the $\leq_1$-interval $(a_{2i-1}^{n+1}, a_{2i+1}^{n+1})$ and $f(a_{2i}^{n+1})$ as the $\leq^2$-least element of the $\leq_2$-interval $(f(a_{2i-1}^{n+1}), f(a_{2i+1}^{n+1}))$. These selections succeed because both orders are dense. It should be clear that the extended sequence a^{n+1} has the same properties specified for the sequence a^n, so this process can be continued to all values of n. Further, it should be clear that the m-th element of the order $\leq^1$ appears in the domain of f by stage m and the m-th element of the order $\leq_2$ appears in the range of f by stage m, so the definition of f succeeds for all values in the domain of $\leq_1$, defines a function which is onto the domain of $\leq_2$, and is clearly a strictly increasing bijection, so an isomorphism.

Definition: A linear order is said to be *complete* iff every subset of the order which is bounded above has a least upper bound.

Observation: The order on $\mathbb{R}$ is complete.

Theorem: A nonempty separable complete dense linear order with no maximum or minimum is isomorphic to the order on $\mathbb{R}$.

Proof: By the theorem above, the order restricted to the countable dense subset is isomorphic to the usual order on $\mathbb{Q}$, from which it follows easily that the entire order is isomorphic to the usual order on $\mathbb{R}$.

3.16.3 Cardinals and Ordinals; Cardinal Successor; The Hartogs and Sierpinski Theorems

For any cardinal $\kappa < |V|$, there are larger cardinals ($|V|$, for instance). Since the natural order on cardinal numbers is a well-ordering, there is a *smallest* cardinal greater than κ. For finite cardinals n, this next largest cardinal is $n+1$, but of course for infinite κ we have $\kappa = \kappa + 1$: we will see below how the "next" cardinal is obtained in the infinite case.

Definition: If $\kappa \neq |V|$ is a cardinal number, we define $\kappa+$ as the least cardinal in the natural order which is greater than κ.

Now we take an apparent digression into the relationships between cardinal and ordinal numbers. Each ordinal α is naturally associated with a particular cardinal:

Definition: Let α be an ordinal number. We define $\texttt{card}(\alpha)$ as $|\texttt{fld}(R)|$ for any $R \in \alpha$ (the choice of R makes no difference).

For each finite cardinal n there is only one ordinal number α such that $\texttt{card}(\alpha) = n$ (usually written n as well). But for any ordinal α such that $\texttt{card}(\alpha)$ is infinite, we find that $\texttt{card}(\alpha + 1) = \texttt{card}(\alpha) + 1 = \texttt{card}(\alpha)$: the `card` operation is far from injective. But there is an ordinal naturally associated with each cardinal as well:

Definition: Let κ be a cardinal. We define $\texttt{init}(\kappa)$ as the smallest ordinal number α such that $\texttt{card}(\alpha) = \kappa$. There is such an ordinal because any set of size κ can be well-ordered; there is a least such ordinal because the natural order on ordinals is a well-ordering.

It is important to note that the T operations on ordinals and cardinals preserve order, addition, multiplication, and exponentiation. Intuitively, this is all true because $T(\kappa)$ is in some external sense the same cardinal as κ and $T(\alpha)$ is in some external sense the same order type as α. The proofs are straightforward but tedious. One has to take into account the fact that cardinal exponentiation is a partial operation (which reflects the fact that there are more cardinals and ordinals in higher types).

We restate and extend our theorems on the fact that the T operation commutes with operations of arithmetic.

Theorem: Let κ and λ be cardinal numbers. Then $T(\kappa) \leq T(\lambda) \leftrightarrow \kappa \leq \lambda$, $T(\kappa) + T(\lambda) = T(\kappa + \lambda)$, $T(\kappa) \cdot T(\lambda) = T(\kappa \cdot \lambda)$, and $T(\kappa^\lambda) = T(\kappa)^{T(\lambda)}$ if the former exists. $T(\kappa+) = T(\kappa)+$.

Theorem: Let α and β be ordinal numbers. Then $T(\alpha) \leq T(\beta) \leftrightarrow \alpha \leq \beta$, $T(\alpha) + T(\beta) = T(\alpha + \beta)$, $T(\alpha) \cdot T(\beta) = T(\alpha \cdot \beta)$, and $T(\alpha^\beta) = T(\alpha)^{T(\beta)}$ (ordinal exponentiation is total).

We now prove a theorem characterizing the way in which $\kappa+$ is obtained from κ when κ is infinite.

Theorem: Let $\kappa = |A| \neq |V|$ be an infinite cardinal. Let Ω_A be the set of order types of well-orderings of subsets of the set A (clearly this does not depend on the choice of the set A). Then $\kappa+ = \mathtt{card}(\sup(\Omega_A))$.

Proof: Let $\gamma = \mathtt{card}(\sup(\Omega_A))$. Since a well-ordering of a set of size γ must be longer than any well-ordering of a subset of A, $\gamma > \kappa$. Now suppose that $\lambda < \gamma$. It follows that $\mathtt{init}(\lambda) < \mathtt{init}(\gamma) = \sup(\Omega_A)$, so a well-ordering of a set of size λ is of the same length as the well-ordering of some subset of a set of size A, so $\lambda \leq \kappa$. Note that the size of the set of ordinals less than $\sup(\Omega_A)$ is $T^2(\gamma)$, so we could also define γ as $T^{-\mathbf{2}}(|\mathtt{seg}_{\leq_\Omega}(\sup(\Omega_A))|)$.

In the absence of Choice the argument above does not work, but there is still something interesting to say.

Definition: For any cardinal $\kappa = |A| \neq |V|$, define Ω_A as the set of order types of well-orderings of subsets of A and $\aleph(\kappa)$ as $\mathtt{card}(\sup(\Omega_A))$.

Observation: The preceding definition is only of interest in the absence of Choice, as otherwise it coincides with $\kappa+$. Note that $\aleph(\kappa)$ is always a cardinal whose elements are well-orderable. Note that for syntactical reasons this use of $\aleph$ is distinguishable from another use to be introduced shortly.

Theorem (not using Choice; Hartogs): For any cardinal κ, $\aleph(\kappa) \not\leq \kappa$.

Proof: Suppose otherwise. Let $\kappa = |A|$. We then have an injection from a set B of order type $\sup(\Omega_A)$ into A. The range of this injection supports a well-ordering of type $\sup(\Omega_A)$. But the range of this injection is a subset of A, so its order type belongs to Ω_A. This is a contradiction.

Theorem (not using Choice; Sierpinski): $\aleph(\kappa) \leq \exp^3(\kappa)$.

Proof: Since we are working in choice-free mathematics, it is advantageous to represent things in different ways. Any well-ordering is represented effectively by the set of its initial segments. We refer to such a representation of an order as a segment-ordering. A segment-ordinal is an equivalence class of segment-ordinals. Notice that a segment-ordinal is three types higher (not two types higher) than the elements of its field. If $A \in \kappa$, observe that the set of segment-ordinals of well-orderings of subsets of A is of cardinality $T^3(\aleph(\kappa))$. Of course a segment-ordinal is a set of sets of sets of elements of A: the collection of sets of sets of sets of elements of A is of cardinality $T^3(\exp^3(\kappa))$. The desired inequality follows.

A related result is $\aleph(\kappa) \leq \exp^2(\kappa^2)$ This is obtained by noting that the usual ordinals of well-orderings of subsets of A are sets of sets of pairs of elements of κ, so $T^2(\aleph(\kappa)) \leq T^2(\exp^2(\kappa^2))$. This is most useful when we know that $\kappa^2 = \kappa$: this is not a theorem of choice-free mathematics, though it is true if elements of κ are well-orderable or if κ is of the form $\exp^4(\lambda)$ for λ infinite (this last because the construction of the Quine pair can then be mimicked in a set of size κ).

3.16.4 Hierarchies of Cardinals; A Disadvantage of Strong Extensionality

We introduce two notations for cardinal numbers.

Definition: Let $\aleph$ be the natural order on infinite cardinals. We then define $\aleph_\alpha$ for ordinals α using the definition of ordinal indexing of the elements of the field of a well-ordering.

Definition: We define ω_α as $\texttt{init}(\aleph_\alpha)$.

Definition: Let $\beth$ be the natural order on cardinal numbers restricted to the smallest set of cardinal numbers which contains $\aleph_0$, is closed under the power set operation, and contains suprema of all of its subsets. We then define $\beth_\alpha$ for ordinals α using the definition of ordinal indexing.

We can now pose one of the notable questions of set theory, dating to the beginnings of the subject. The first infinite cardinal is $\aleph_0$. We know by

Cantor's Theorem that $|\mathcal{P}(\mathbb{N})| > |\iota``\mathbb{N}| = \aleph_0$. We also note that $|\mathcal{P}(\mathbb{N})| = \beth_1$. We know by definition of cardinal successor that $\aleph_1 = \aleph_0+ > \aleph_0$. We know by the observation following the theorem above that $\aleph_1$ is the number of finite and countable ordinals (which is easily shown to be the same as the number of countable ordinals). The question that arises is the status of

*__Cantor's Continuum Hypothesis:__ $\beth_1 = \aleph_1$? Are there more subsets of the natural numbers than countable order types?

It is called the Continuum Hypothesis because Cantor also knew (as we will find in the next section) that $\beth_1$ is not only the cardinality of the the set of subsets of the natural numbers but also the cardinality of the set of real numbers, or the number of points on a line (the cardinality of the continuum). For this reason $\beth_1$ is also called c (for "continuum").

A related assertion (which is again a hypothesis not a theorem) is

*__Generalized Continuum Hypothesis (GCH):__ $\aleph_\alpha = \beth_\alpha$ for all ordinals for which $\aleph_\alpha$ is defined.

A further question is how far the $\aleph_\alpha$'s or $\beth_\alpha$'s continue. These notations are definitely undefined for sufficiently large ordinals α (neither is defined for $\mathtt{ot}(\aleph)$, by a simple consideration of how ordinal indexing is defined). We cannot prove in this system that $\aleph_\omega$ is defined or even that $\aleph_n$ exists for each natural number n. It is true that $\beth_n$ exists among the cardinals of type n sets for each n, but there is a kind of pun going on here. It is also true that the sequences of $\aleph$'s and $\beth$'s get longer in higher types. Suppose $|V| = \aleph_\alpha$ (there will be such an α). It follows that $T(|V|) = \aleph_{T(\alpha)}$ in the next higher type, so the strictly larger cardinal $|\mathcal{P}(V)| \geq \aleph_{T(\alpha)+1}$, so the sequence is extended in length by at least one. A similar argument for the $\beth_\alpha$'s is slightly more involved.

With strong extensionality there is a much stronger restriction. Suppose that the cardinality of type n is $\aleph_\alpha$. It follows that the largest $\beth_\beta$ which is the cardinality of a type $n+1$ set has $\beta \leq \alpha$. Further, it follows that the largest $\beth_\alpha$ which is the cardinality of a type $n+2$ set has β no greater than $T(\alpha) + 2$. Iteration of this observation (and the natural identification of ordinals of different types via the T operation) allow us to say somewhat loosely that there can be no $\beth_\beta$ in any type with $\beta \geq \alpha + \omega$. The reason for this restriction is that there is a definable bound on the size of each type $n+1$ in terms of the size of type n.

This gives a concrete motivation for the form of the axiom of extensionality that we have chosen to use. We do not want the size of mathematical structures that we can consider to be strongly bounded by the size of type 0.

With weak extensionality we can cause much larger $\beth$ numbers to exist because we can assume that each type $n+1$ is much much larger than the power set of type n (a sufficiently large set of urelements is added to support whatever construction we are considering). A strong assumption which suggests itself is that we can iterate the cardinal exponentiation operation on cardinals of sets of type n objects along any well-ordering of type n objects (for each type n). This would give existence of $\beth_{T^{\mathbf{2}}(\alpha)}$ for each ordinal α.

It is useful to note that if we use the convention of type-free cardinal and ordinal variables outlined above, we can treat the exponential operation on cardinals as total. This is achieved in the underlying translation to typed language by providing that we work in a type higher than that of any variable appearing in an exponentiation: the exponential κ^λ is then in effect read as $T(\kappa)^{T(\lambda)}$, which is always defined.

This means that we can in effect say "For every cardinal there is a larger cardinal" and "For every ordinal there is a larger ordinal". $(\forall\kappa.(\exists\lambda.\lambda > \kappa))$ do not make sense under the convention, because we have not bounded the quantifiers. But $(\forall\kappa < \mu.(\exists\lambda < 2^\mu.\kappa < \lambda))$ is true (for any specific μ, with the convention ensuring that we work in a type where 2^μ exists), and expresses the desired thought.

3.17 Sets of Reals

topological stuff?

3.18 Complex Type Theories

complicated type theories and how they can be represented in TSTU; Curry-Howard isomorphism stuff, perhaps.

3.19 Infinite Indexed Families; König's Theorem

3.20 Partition Relations

We begin this section by stating an obvious

Theorem: If X is an infinite set, A is a finite set, and $f : X \to A$, then there is $a \in A$ such that $f^{-1}\text{“}\{a\}$ is an infinite subset of X.

Proof: The preimages of individual elements of A under f are a disjoint finite family of sets covering X. The sum of their cardinalities is the cardinality of X. If all of them had finite cardinality, this sum (the cardinality of X) would be finite. But the cardinality of X is infinite by assumption.

The first major theorem of this section is a generalization of this.

Definition: If X is a set and κ is a cardinal, we define $[X]^\kappa$ as $\mathcal{P}(X) \cap \kappa$, the set of all subsets of X of size κ.

Definition: If X is a set, n is a natural number, A is a finite set, and $f : [X]^n \to A$, we say that H is a homogeneous set for f iff $H \subseteq X$ and $|f\text{“}[H]^n| = 1$.

Theorem (Ramsey): If X is an infinite set, n is a natural number, A is a finite set, and $f : [X]^n \to A$, there is an infinite homogeneous set H for f.

Proof: For $n = 1$, the result follows immediately from the first theorem of this section.

Assume that the theorem is true for $n = k$ and show that it follows for $n = k + 1$. Let X be an infinite set, A a finite set, and $f : [X]^{k+1} \to A$ a function. Our goal is to show that there is an infinite homogeneous set H for f.

We define a tree T_f. We well-order X and we assume that $u\, T_f\, v$ is defined for all $u, v \leq x$. We define $x\, T_f\, u$ as false for $u < x$. We define $y\, T_f\, x$, for $y < x$, as true iff for all k-element subsets K of $\mathtt{seg}_{T_f}(y)$, $f(K \cup \{y\}) = f(K \cup \{x\})$. $\leq_f$ is a tree because the order on any segment in $\leq_f$ agrees with the underlying well-ordering of X.

We introduce some terminology useful in the context of trees. The *level* of an element x of the field of a tree $\leq_T$ is the order type of $\mathtt{seg}_{\leq_T}(x)$. The *branching* of the tree at an element x of its field is the cardinality of the set of all y such that x is the maximal element of the segment determined by y in the tree. Such elements y are called successors of x in the tree.

In the tree $\leq_f$, the branching will be finite at any element of the field of the tree at finite level. There will be nontrivial branching above an element y just in case there are elements z, w such that for any k-element subset K of $\mathtt{seg}_{\leq_f}(y)$, $f(K \cup \{y\}) = f(K \cup \{z\}) = f(K \cup \{w\})$, but for some $k-1$-element subset A, $f(A \cup \{y, z\}) \neq f(A \cup \{y, w\})$. A possible new branch above y is determined by the assignment of a value under f to each $f(A \cup \{y, z\})$, where z is the next element of the branch. Since there are finitely many subsets of the segment determined by y (since its level is finite) and finitely many values in A, the branching at each element of finite level is finite. One can further prove that if the branching at each element of a finite level is finite, each finite level is a finite set. It follows that some element of each finite level is dominated by infinitely many members of X in the tree order, and further that if some element of finite level is dominated by infinitely many elements of X, it has a successor that is dominated by infinitely many elements of X. From this it follows that we can construct a branch of the tree with the property that each of its elements of finite level is dominated by infinitely many elements of X (so it has elements of all finite levels and is infinite).

Any branch B in the tree $\leq_f$ has the property that if $b_1 \leq_f b_2 \leq_f \ldots b_k \leq_f b_{k+1} \leq_f c$, that $f(\{b_1, b_2, \ldots, b_k, b_{k+1}\}) = f(\{b_1, b_2, \ldots, b_k, c\})$: the value at a $k+1$-element subset of the branch is not changed if the top element of the set is changed. Thus we can define a new function $f^* : [B]^k \to A$ by $f(\{b_1, b_2, \ldots, b_k\}) = f(\{b_1, b_2, \ldots, b_k, c\})$ for any $c \geq_f b_k$. Now let B be the infinite branch whose existence was shown above. By inductive hypothesis, there is an infinite homogeneous set H for B with respect to f^*, which will also be an infinite homogeneous set for f. This completes the proof.

Ramsey theorem and Erdös-Rado theorem: not part of the main agenda, used for model theory of alternative set theories later.

The Schmerl partition relations, needed for theory of NFUA.

3.21 Large Cardinals

inaccessibles, Mahlos, weakly compact and measurables explained. This is prerequisite knowledge for the model theory of strong extensions of *NFU*; it can also be used to talk about model theory of *ZFC*.

3.22 Pictures of Sets: the Theory of Isomorphism Types of Well Founded Extensional Relations

In this section, we show how the type theory we are working in can naturally motivate a development of the untyped set theory which is more often used, as the theory of a quite natural class of mathematical structures which has its own intrinsic interest.

3.22.1 Coding Sets in Relations

We consider the possibility that a set relation R may be used to represent the membership "relation" $\in$. Toward this end, we introduce some definitions.

Definition: Let R be a relation. We say that an element x of $\mathtt{fld}(R)$ *codes* the set $R^{-1}\text{“}\{x\} = \{y \mid y\,R\,x\}$ relative to R. (if the relation is understood in the context we may just say that the element x codes the given set).

The Definition ensures that a given domain element codes just one subset of the field of the relation, but we would also like it to be the case that a given set is coded by no more than one domain element.

Definition: A relation R is said to be *[weakly] extensional* iff for all x and y in the field of R, if $R^{-1}\text{“}\{x\} = R^{-1}\text{“}\{y\}$ then [either $R^{-1}\text{“}\{x\} = R^{-1}\text{“}\{y\} = \emptyset$] or $x = y$.

A weakly extensional relation leaves open the possibility of coding a theory of sets with distinct urelements, such as are allowed to exist in our type theory: there may be many distinct R-minimal objects if R is weakly extensional, but only one if R is extensional.

Because we are working with set relations, we perforce are at least tempted to use untyped language. For example, we can ask the question whether there is a code for the set $\{x \in \mathtt{fld}(R) \mid \neg x\,R\,x\}$ relative to the relation R. The argument for Russell's paradox shows us that there cannot be such a code (though the set certainly exists). In our type theory we cannot even ask the question which leads to Russell's paradox.

A notion which is difficult (though not entirely impossible) to develop in type theory is the notion of the collection of elements of a set, elements of its elements, elements of elements of its elements, and so forth (a kind of downward closure). In the theory of coded sets this is straightforward.

Definition: Let R be a relation (we do not require it to be extensional). Let x be an element of the field of R. We define the *component* of x determined by R as $R \cap D_x(R)^2$, where $D_x(R)$ is the minimal subset of the field of R which contains x as an element and contains $R^{-1}\text{“}\{y\}$ as a subset for each of its elements y. We denote the component of x determined by R by $C_x(R)$.

Theorem: Let R^* be the minimal reflexive, transitive relation which includes R. Then $C_y(R)$ is $R \cap \{x \mid x\, R^*\, y\}^2$.

Proof: $x \in D_x(R)$ is obvious. Suppose $x \in D_y(R)$ and $y \in D_z(R)$. Any set which contains z as an element and which includes $R^{-1}\text{“}\{u\}$ as a subset for each of its elements u must contain y (by definition of $D_z(R)$ and the fact that $y \in D_z(R)$) and so further must contain x (by definition of $D_y(R)$ and the fact that $x \in D_y(R)$) so we have shown that $x \in D_z(R)$. Thus the relation $x\, S\, y$ defined as $x \in D_y(R)$ is reflexive and transitive, so $x\, R^*\, y$ implies $x \in D_y(R)$. Now observe that $\{y \mid y\, R^*\, x\}$ contains x and includes the preimage under R of any of its elements, so must be included in $D_y(R)$. We now see that the field $D_y(R)$ of the component $C_y(R)$ is precisely $\{x \mid x\, R^* y\}$, from which the result follows.

There is a notion of isomorphism appropriate to weakly extensional relations.

Definition: If R and S are weakly extensional relations, we say that f is a *membership-isomorphism* from R to S if f is a bijection from the field of R to the field of S such that $x\, R\, y \leftrightarrow f(x)\, S\, f(y)$ and in addition if $R^{-1}\text{“}\{x\} = S^{-1}(f(x)) = \emptyset$ it also follows that $x = f(x)$.

We impose a further condition on relations which we regard as simulating the membership relation, for which we need to supply a motivation.

Definition: A *[weak] membership diagram* is a well-founded [weakly] extensional relation.

Theorem: If R is well-founded, so is $R^* - [=]$.

Proof: Suppose A is a nonempty subset of $\texttt{fld}(R^* - [=])$ with no $(R^* - [=])$-minimal element. Certainly A is a nonempty subset of $\texttt{fld}(R)$. Let a be an R-minimal element of A. There must be $b \neq a$ such that $b\, R^*\, a$

(since there is no $(R^* - [=])$-minimal element). But from $b\,R^*\,a$, it is easy to deduce $(\exists x.x\,R\,a)$, which is a contradiction.

The effect of the well-foundedness restriction is to ensure that if R and S are membership diagrams and f is a membership-isomorphism from R to S, we can be certain that x with respect to R and $f(x)$ with respect to S always "represent precisely the same set". It is somewhat difficult to say precisely what is meant by this (since we do not yet have an independent understanding of untyped set theory), but a definite result which we can state is that the membership-isomorphism f is *unique*: there can be no other membership-isomorphism from R to S. Suppose there was another such membership-isomorphism g. There would be an R-minimal x in the domain of R such that $f(x) \neq g(x)$. If the R-preimage of x were empty, then so would be the S-preimages of $f(x)$ and $g(x)$, but further we would have $x = f(x) = g(x)$, contradicting the choice of x as a counterexample. If the R-preimage of x were a nonempty set A, then the S-preimage of $f(x)$ would be $f\text{“}A$ and the S-preimage of $g(x)$ would be $g\text{“}A$. But by minimality of x, $f\text{“}A = g\text{“}A$, so by extensionality of S, $f(x) = g(x)$, contradicting the choice of x as a counterexample.

The informal argument that each element of x designates the same set relative to R that is designated by $f(x)$ with respect to S has the same form, but has an essential vagueness dictated by the fact that we are not actually previously acquainted with the domain of sets being designated. If the R-preimage of x is empty, then $x = f(x)$: the two objects represent the same atom. If the R-preimage of x is a nonempty set A, then the S-preimage of $f(x)$ is $f\text{“}A$. x designates (with respect to R) the collection of things designated by elements of A with respect to R. By the minimality hypothesis, the things designated with respect to S by elements of $f\text{“}A$ are the same: so $f(x)$ designates the same collection with respect to S that x designates with respect to R.

Further, if we are working with relations that are extensional rather than weakly extensional, the argument above works with isomorphism in place of membership-isomorphism.

General well-founded relations can be "collapsed" to well-founded [weakly] extensional relations in a suitable sense.

Theorem: Let R be a well-founded relation. Then there is a uniquely determined equivalence relation $\sim$ on $\texttt{fld}(R)$ with the following property (in

which we use the notation $[x]$ for $[x]_\sim$): the relation $R_\sim = \{\langle [x], [y]\rangle \mid x\,R\,y\}$ is [weakly] extensional and for each $[x]$ we have the set of its $R_\sim$-preimages exactly the set of $[y]$ such that $y\,R\,x$.

Proof: Let x be minimal in R such that $C_x(R)$ does not have this property. (Clearly if there is no such x, then the unions of uniquely determined equivalence relations on all $C_x(R)$'s with the indicated property will give such an equivalence relation on R.) Each $C_y(R)$ for $y\,R\,x$ will support such a unique equivalence relation, if it is nonempty. We define the desired equivalence relation on $C_x(R)$, contrary to hypothesis. The top x is equivalent only to itself. All R-preimages of x which have empty R-preimage are either equivalent only to themselves (if we are working with membership-isomorphism) or equivalent to all such preimages (if we are working with isomorphism). Each other element y of $C_x(R)$ has an associated equivalence relation $\sim$ and relation $C_y(R)_\sim$: define $y \sim z$ as holding if and only if $C_y(R)_\sim$ is [membership]-isomorphic to $C_z(R)_\sim$. By hypothesis the restriction of the equivalence relation to each proper component is unique. Extensionality (and the known uniqueness of isomorphisms between well-founded [weakly] extensional relations] leaves us no freedom of choice with respect to defining the equivalence relation between elements of different components. So the equivalence relation obtained is unique.

To see why non-well-founded "membership diagrams" are problematic, consider a diagram containing two elements x and y, each related just to itself. This codes two sets, each of which is its own sole element. Consider another diagram containing two elements u and v, each related just to itself. Either of the two bijections between the fields of these relations is a membership-isomorphism (and indeed an isomorphism) between the relations: there is no way to determine whether x is to be identified with u or with v.

It should be noted that non-well-founded "membership diagrams" are *merely* problematic, not impossible. Interesting untyped theories can be developed in which there are objects which are their own sole elements (and in which there can be many such objects), and in fact we will have occasion to see this later. Indeed, arbitarily complex failures of well-foundedness of the membership relation are possible and worthy of study.

3.22.2 Passing to Isomorphism Types

The advantage of restricting ourselves to well-founded [weak] membership diagrams is that for any element x of the field of a well-founded membership diagram R, the intended reference of x is in effect fixed by the [membership-]isomorphism type of the component $C_x(R)$. We can then view the [membership-]isomorphism types of components of diagrams as the actual objects under study. When studying weak membership diagrams, there is an element of arbitrariness in the choice of atoms, though it is sometimes useful to have atoms in untyped set theory. The isomorphism types of well-founded extensional relations will be our principal study, and we will see that they correspond precisely to the objects of the usual untyped set theory, though without strong assumptions we will not see the *entire* universe of the usual set theory [in whatever sense this is possible].

Observation: If a [weak] set diagram R is equal to $C_x(R)$ and to $C_y(R)$ where x and y belong to the field of R, then $x = y$. This condition implies $x\,R^*\,y$ and $y\,R^*x$. Since $R^* - [=]$ is well-founded, an $(R^* - [=])$-minimal element of $\{x, y\}$ must be equal to both x and y, so $x = y$.

Definition: A *weak set diagram* is a weak membership diagram which is equal to one of its components (and thus must be nonempty). A *set diagram* is a membership diagram which is either empty or equal to one of its components. A *top* of a [weak] set diagram is either the unique x such that the diagram is its own component determined by x or (in case the diagram is empty) any object whatsoever. A *[weak] set picture* is the [membership-]isomorphism class of a [weak] set diagram [or a double singleton (representing an atom)]. The set of all set pictures is called Z. The set of all weak set pictures whose elements have atoms restricted to a set A is called $Z[A]$ (this last will contain only double singletons of elements of A; of course $Z[V]$ contains all weak set pictures).

Definition: For any [weak] set diagram R with top t, we define an *immediate component* of R as a component $C_x(R)$ such that $x\,R\,t$. Note that the empty set diagram has no immediate components, but may occur as an immediate component of a set diagram if the $x\,R\,t$ happens to have empty R-preimage: the handling of elementless objects in weak set diagrams is seen below. For set pictures ρ and σ, we define $\rho\,E\,\sigma$ as holding iff there are $R \in \rho$ and $S \in \sigma$ such that R is an immediate

component of S. For weak set pictures ρ and σ, we define $\rho\, E\, \sigma$ as holding iff there are $R \in \rho$ and $S \in \sigma$ such that R is an immediate component of S, or ρ is a double singleton $\{\{x\}\}$, and σ has an element S with top t such that $x\, S\, t$ and the S-preimage of x is empty (this handles atoms). It is important to note that no double singleton is a membership-isomorphism class of weak set diagrams, so there is no conflict between the two parts of the definition of E on weak set pictures (the double singleton of the empty set is a set picture, and the sole elementless object in the "set theory" implemented using set diagrams).

Theorem: E is a membership diagram (on Z or $Z[A]$).

Proof: We need to show that E is [weakly] extensional and that E is well-founded. Suppose that ρ and σ are [weak] set pictures and $E^{-1}\text{“}\{\rho\} = E^{-1}\text{“}\{\sigma\}$. This means that each immediate component of any $R \in \rho$ is isomorphic to some immediate component of any $S \in \sigma$ and vice versa. [In the weak case, any preimage of the top of R which has empty R-preimage is identical to some preimage of the top of S which has empty S-preimage, and vice versa]. There is a unique isomorphism from the field of each immediate component of R to a uniquely determined immediate component of S (because no two distinct immediate components can be isomorphic). Any two of these isomorphisms will agree on any common element of their domains. It follows that the union of these isomorphisms, taken together with the pair whose first projection is the top of R and whose second projection is the top of S, yields a [membership-]isomorphism from R to S, so $\rho = \sigma$. [The fact that it is a membership-isomorphism in the weak case follows from the bracketed complete sentence above: elements of $E^{-1}\text{“}\{\rho\}$ and $E^{-1}\text{“}\{\sigma\}$ which are double singletons each correspond to identical elements of the other, and this allows one to define the isomorphism so that it fixes all elements with empty E-preimage]. We have shown that E is [weakly] extensional.

Suppose that A is a nonempty subset of the field of E. Let ρ be an element of A and let $R \in \rho$. [If R is a double singleton, R is E-minimal and we are done.] Define A_R as the intersection of A with the set of isomorphism types of components $C_x(R)$ [and double singletons of R-minimal elements of the field of R]. There will be a minimal x such that the isomorphism type of $C_x(R)$ belongs to A [or there will be a

double singleton which belongs to A]; the isomorphism class of $C_x(R)$ [or the double singleton] will be an E-minimal element of A_R, and so an E-minimal element of A.

Observation: Note that E is two types higher than the [weak] membership diagrams R with which we started. If x in the field of R is at type k, then R itself is at type $k+1$, the [membership-]isomorphism class of R is at type $k+2$, and E is at type $k+3$. We see that E is two types higher than the arbitrary membership diagrams with which we started. E is a kind of universal membership diagram, but this type differential will allow us to completely naturally evade any supposed paradoxical consequences of this universality. The situation here is analogous to that for ordinals: the well-ordering on all ordinals is a kind of universal well-ordering – it contains not a suborder isomorphic to each well-ordering R but a suborder isomorphic to the double singleton image $R^{\iota^{\mathbf{2}}}$ of each well-ordering R. It is also worth noting that strict well-orderings with maxima (and the empty strict well-ordering) are well-founded extensional relations, so there are elements of Z (or $Z[A]$) naturally related to the ordinal numbers (and indeed these correspond precisely to the objects (the *von Neumann ordinals*) which are normally taken to be the ordinal numbers in the usual set theory). One must observe though that a nonzero ordinal α is implemented in untyped set theory by the isomorphism class of the strict well-ordering derived from a well-ordering of order type $\alpha+1$.

There is a type-shifting operation T on [weak] set pictures analogous to the operations on cardinals and ordinals.

Definition: For any [weak] set diagram R, define R^ι as usual: this will still be a [weak] set diagram. Let ρ be the [membership-]isomorphism class of R: then $T(\rho)$ is defined as the [membership]-isomorphism class of R^ι, and it is straightforward to show that the specific choice of an element R of ρ has no effect on the definition of $T(\rho)$. Notice that in the case of weak set diagrams, atoms are replaced by their singletons as we pass up one type. [Define $T(\{\{x\}\})$ as $\{\{\{x\}\}\}$].

Theorem: For all [weak] set pictures ρ and σ, $\rho\, E\, \sigma \leftrightarrow T(\rho)\, E\, T(\sigma)$.

Proof: This follows directly from the precise parallelism of the structure of $S \in \sigma$ with the structure of $S^\iota \in T(\sigma)$. If $\rho\, E\, \sigma$, any $S \in \sigma$ has

an immediate component $R \in \rho$, so belonging to ρ: it is immediate that $S^\iota \in T(\sigma)$ has an immediate component R^ι belonging to $T(\rho)$, so $T(\rho) \in T(\sigma)$. Suppose $T(\rho) \in T(\sigma)$. Then we can choose an element of $T(\sigma)$ of the form S^ι where $S \in \sigma$, which will have an immediate component $R^\iota \in T(\rho)$ (any component of S^ι is obviously a relation singleton image), from which we discover $R \in \rho$, so $\rho \in \sigma$. [If the top of $S \in \sigma$ has an immediate preimage x with empty S-preimage, and $\rho = \{\{x\}\}$, then the top of S^ι has an immediate preimage $\{x\}$, so $\{\{\{x\}\}\} = T(\rho)\,E\,T(\sigma)$ in this case as well; if $T(\rho) \in T(\sigma)$ where $\rho = \{\{x\}\}$, the top of $S^\iota \in T(\sigma)$ has an immediate preimage $\{x\}$ with empty S^ι-preimage [recall that we can without loss of generality choose an element of σ of the form S^ι], we see that the top of $S \in \sigma$ has the preimage x with empty S-preimage, so $\{\{x\}\} = \rho\,E\,\sigma$].

Theorem: For each $\rho \in Z$ $[Z[A]]$ we have $C_\rho(E) \in T^{\mathbf{2}}(\rho)$.

Proof: Let $R \in \rho$. Define ρ_x as the isomorphism type of $C_x(R)$ for $x \in \mathtt{fld}(R)$ [or as $\{\{x\}\}$ if x is R-minimal.] The ρ_x's are exactly the elements of $D_\rho(E)$, $\rho_x\,E\,\rho_y$ iff $x\,R\,y$, but ρ_x is two types higher than x, so we can define a [membership-]isomorphism sending each $\{\{x\}\}$ to ρ_x, witnessing the desired relation between $R^{\iota^{\mathbf{2}}}$ and $C_\rho(E)$.

Theorem (using Choice): Every subset of $T\text{“}Z[A]$ is coded in E. Every subset of $T\text{“}Z$ is coded in E.

Proof: Let B be an arbitrary subset of $T\text{“}Z$ $[T\text{“}Z[A]]$. Each element of B is of the form $T(\rho)$ We transform each $R^\iota \in T(\rho)$ for each $\rho \in B$ to a different R' still belonging to $T(\rho)$: $R' = \{\langle\langle\{x\}, R\rangle, \langle\{y\}, R\rangle\rangle \mid x\,R\,y\}$. The collection of relations R' is pairwise disjoint, so we can take their union and adjoin all pairs $\langle t, T\rangle$ as new elements, where t is the top of one of the R''s and T is a fixed new top element (any pair whose second projection does not belong to B will do). The resulting relation is well-founded and has immediate components of exactly the right isomorphism classes, but it is not extensional. By the theorem proved above on collapsing well-founded relations to well-founded [weakly] extensional relations, we can define an equivalence relation on its field and replace each element of the field by a representative of its equivalence class taken from a fixed choice set in such a way as to obtain a

[weak] set diagram which has immediate components with isomorphism classes which are all and only the elements of B.

Theorem (not using Choice): Every subset of $T^{\mathbf{2}}\text{“}Z[A]$ is coded in E. Every subset of $T^{\mathbf{2}}\text{“}Z$ is coded in E.

Proof: Let B be a subset of $T^{\mathbf{2}}\text{“}Z$ $[T^{\mathbf{2}}\text{“}Z[A]]$. Each element $T^{\mathbf{2}}(\rho)$ of B has a *canonical* representative, namely $C_\rho(E)$. These relations all agree on shared members of their domains (since they are all subsets of E). Add a new top element T and add all pairs $\langle \rho, T\rangle$ for $T^{\mathbf{2}}(\rho) \in B$ as elements to their union to obtain a relation with the correct isomorphism classes of immediate components.

Observation: The membership diagram E in higher types faithfully reproduces the membership diagrams in the E relations in lower types. Moreover, the E relation in higher types is *complete* in an obvious sense on its copy of the domains of the E relation of lower types: it codes all subsets of the domains at lower types, whereas a specific E relation cannot code all subsets of *its own* domain. For example, a specific relation E cannot code its own field $Z = \mathtt{fld}(E)$, because it is a well-founded relation (a code v for the entire field of E would satisfy $v\,E\,v$). But $T\text{“}\mathtt{fld}(E)$ is coded (in E of a higher type) from which we can see that more sets are coded in higher types.

3.22.3 The Hierarchy of Ranks of Set Pictures

We introduce the analogue here of the cumulative hierarchy of sets in the usual set theory – without atoms. From this point on we restrict ourselves to membership diagrams, though the results for weak membership diagrams are quite similar.

Definition: For any set $A \subseteq \mathtt{fld}(E)$, we define $P(A)$ as the set of elements of $\mathtt{fld}(E)$ which code subsets of A. We say that the subset A is *complete* if $P(A)$ contains codes for all subsets of A. Notice that $P(A)$ has the same type as A.

Definition: We define the set of *ranks in* E as the intersection of all sets H such that $\emptyset \in H$, $(\forall h \in H.P(h) \in H)$, and $(\forall A \subseteq H.\bigcup A \in H)$.

Theorem: The set of ranks itself contains $\emptyset$, is closed under P and closed under unions of sets of ranks. The ranks in E are well-ordered by inclusion.

Theorem: $\mathtt{fld}(E)$ is a rank.

Definition: Let $\mathbb{E}$ denote the inclusion order on ranks in E. Then $\mathbb{E}_\alpha$ is a general notation for ranks using our convention on ordinal indexing.

Definition: Let γ be the ordinal such that $\mathbb{E}_\gamma$ is the first incomplete rank.

Theorem: $\mathbb{E}_{\omega+n}$ is a complete rank in a high enough type for each familiar natural number n.

Theorem: $|\mathbb{E}_{\omega+\alpha}| = \beth_\alpha$ if $\mathbb{E}_{\omega+\alpha}$ is complete.

The ranks code an iterative process for constructing sets by iterating the "power set" construction which may go through stages indexed by infinite ordinals. This is reminiscent of how the world of our type theory is constructed, except that we lack the ability (or indeed the need) to pass to transfinite levels.[2].

The set pictures are isomorphism classes supporting a T operation, so we can introduce type free variables ranging over set pictures using the conventions introduced above. Each set picture variable needs to be restricted to some definite type, which can be viewed as restriction of the variable to some set of set picture variables (in higher types) which can in turn be viewed as restriction of the variable to the preimage under E of some set picture (if we go up one more type so that all elements of the original type are images under T so we have completeness). Just as we represented the bounding of ordinal variables in types as bounding in the segment determined by an ordinal variable, we can represent the bounding of set picture variables in types or sets within types as a bounding in the preimage of a set picture under E.

The self-contained theory of set pictures thus obtained is an untyped set theory with E as its membership.

We outline the proofs of some important theorems of this untyped theory.

[2]We will explore further the question as to whether type theory suffers from the lack of transfinite levels. But notice that we are able to discuss the transfinite levels of the cumulative hierarchy in type theory here, and the possible presence of urelements means that the hierarchy will not necessarily be truncated at any definite point as it would be in a strongly extensional development of type theory

Theorem: For every set picture σ and every formula ϕ, there is a set picture τ such that $(\forall \rho\, E\, \tau . \rho\, E\, \sigma)$ and $(\forall \rho \in \sigma . \rho \in \tau \leftrightarrow \phi)$.

Proof: Our conventions ensure that we work in a type where $\sigma = T(\sigma')$ for some σ', and the result then follows from theorems given above: the image under T of any set of set pictures is coded.

Theorem: For every set picture σ, the set of all codes of subsets of the preimage of σ under E is coded.

Proof: Just as with the result that cardinal exponentiation is total in the untyped theory of cardinals, this is achieved by clever definition of our conventions. We stipulate that if any set picture σ is mentioned, we work in a type high enough that $\sigma = T(\sigma')$ for some σ'. This ensures that any subcollection of the preimage of σ under E is coded (the burden of the previous theorem) and is further itself also an image under T, so the collection of all these subsets is also coded (though it is not necessarily an image under T). Note that if we further mention this set (for a specific σ) we bump ourselves into a yet higher type (so we can iterate this "power set" operation any concrete finite number of times).

4 Untyped theory of sets

In this section we introduce the usual untyped set theories (Zermelo set theory and the stronger *ZFC*) and relate them to type theory. We present the view that untyped set theory can be interpreted as the theory of set pictures (isomorphism types of certain well-founded extensional relations), which should already be suggested by the treatment at the end of the previous section.

Further, we strongly criticize the idea that the axioms of Zermelo set theory are somehow essentially *ad hoc*, as is often suggested (this is stated with great confidence so often as to be cliché). There are some odd features of the earliest form of the axioms, which reflect the fact that they appear early in the process of understanding what can be done with set theory, but Zermelo set theory is very close to being exactly the abstract theory of set pictures, and this is not *ad hoc*.

In untyped set theory there is only one kind of object – sets. There may also be atoms if extensionality is weakened to allow them but they will not be an essentially different sort (type) of object. Though this may seem to be quite a different kind of theory, we will see that the usual untyped set theory is not so distantly related to the typed theory of sets we have developed as you might think.

4.1 The original system of Zermelo

The first modern axiomatic system of set theory was proposed by Zermelo in 1908. It is even older than the first publication of the famous *Principia Mathematica* of Russell and Whitehead, though not as old as Russell's first proposal of the theory of types.

The axioms differ somewhat from those in modern treatments. In this theory, we have primitive predicates of membership and equality, and all objects are of the same sort (there are no type restrictions in our language).

Extensionality: Sets with the same elements are the same. It appears that Zermelo may allow atoms (non-sets) in his original formulation, but we will assume here that all objects are sets.

Elementary Sets: The empty set $\emptyset$ exists. For any objects x and y, $\{x\}$ and $\{x, y\}$ are sets.

Separation: For any property $\phi[x]$ and set A, the set $\{x \in A \mid \phi[x]\}$ exists.

Power Set: For any set A, the set $\{B \mid B \subseteq A\}$ exists. The definition of $A \subseteq B$ is the usual one.

Union: For any set A, the set $\bigcup A = \{x \mid (\exists y \in A.x \in y)\}$ exists.

Infinity: There is a set I such that $\emptyset \in I$ and $(\forall x.x \in I \rightarrow \{x\} \in I)$.

Choice: Any pairwise disjoint collection of nonempty sets has a choice set.

We give some discussion of the axioms.

We will assume strong extensionality (objects with the same elements are equal), as is now usual, but note that Zermelo was prepared to allow non-sets with no elements. In type theory we use a weaker form of extensionality because the strong form of extensionality imposes a strong restriction on how

large the universe of our type theory can be; in Zermelo set theory this is not the case.

The axiom of elementary sets is more complicated than is necessary. The separate provision of the singleton set is not made in the modern treatment, as $\{x\} = \{x, x\}$ exists if we merely assert the existence of unordered pairs, and Separation and Infinity together imply the existence of the empty set ($\emptyset = \{x \in I \mid x \neq x\}$).

Zermelo did not know that the ordered pair could be defined by $\langle x, y\rangle = \{\{x\}, \{x, y\}\}$, but note that the ordered pair (now in the Kuratowski form) is provided by the axiom of elementary sets.

The axiom of separation does not appear to imply any paradoxes. We attempt the Russell argument: define $R_A = \{x \in A \mid x \notin x\}$. Observe that $R_A \in R_A \leftrightarrow R_A \in A \wedge R_A \notin R_A$. This would only lead to contradiction if $R_A \in A$, so we conclude $R_A \notin A$, whence we conclude that there is no universal set (for every set A we have specified a set R_A which cannot belong to it).

The axiom of power set and the axiom of union define familiar constructions. Note that $x \cup y$ can be defined as $\bigcup\{x, y\}$. $x \cap y = \{z \in x \mid z \in y\}$ and $x - y = \{z \in x \mid z \notin y\}$ are provided by Separation alone. Complements do not exist for any set. The cartesian product $A \times B$ is definable as $\{c \in \mathcal{P}^{\mathbf{2}}(A \cup B) \mid (\exists ab.a \in A \wedge b \in B \wedge c = \langle a, b\rangle)\}$.

For Zermelo the natural numbers were coded as iterated singletons of the empty set. Given the set I provided by the axiom of infinity, and terming sets containing $\emptyset$ and closed under singleton "Zermelo-inductive" sets, we can define $\mathbb{N}$ as $\{n \in I \mid n$ belongs to all Zermelo-inductive sets$\}$.

In a modern treatment, the von Neumann successor x^+ is defined as $x \subseteq \{x\}$, and the axiom of infinity asserts that there is a set which contains the empty set and is closed under the von Neumann successor operation. It is interesting to observe that neither form of the axiom of infinity implies the other in the presence of the other Zermelo axioms (though they are equivalent in the presence of the axiom of replacement).

It is remarkable that in spite of the fact that Zermelo did not know how to code the general theory of relations and functions into set theory (lacking an ordered pair definition) he was able to prove the Well-Ordering Theorem from the Axiom of Choice in his 1908 paper. Some day I have to look at how he did it!

The axioms of Foundation and Replacement which complete the modern set theory *ZFC* were later developments.

We describe a minimal model of Zermelo set theory. The domain of this model is the union of the sets $\mathcal{P}^{\mathbf{i}}(\mathbb{N})$. It is important to note that the Zermelo axioms give us no warrant for believing that this sequence of sets make up a set. Extensionality certainly holds in this structure. The empty set belongs to $\mathbb{N}$, so is certainly found in this structure. It is useful at this point to note that $\mathbb{N} \subseteq \mathcal{P}(\mathbb{N})$ (each Zermelo natural number is a set of Zermelo natural numbers, 0 being the empty set and $n+1$ being $\{n\}$); since $A \subseteq B$ obviously implies $\mathcal{P}(A) \subseteq \mathcal{P}(B)$, we have (by repeated application) $\mathcal{P}^{\mathbf{i}}(\mathbb{N}) \subseteq \mathcal{P}^{i+1}(\mathbb{N})$ and so $\mathcal{P}^{\mathbf{i}}(\mathbb{N}) \subseteq \mathcal{P}^{j}(\mathbb{N})$ if $i \leq j$. The iterated power sets of the set of natural numbers whose union is our structure are nested. For any x and y in the structure, there are m and n such that $x \in \mathcal{P}^{\mathbf{m}}(\mathbb{N})$ and $y \in \mathcal{P}^{\mathbf{n}}(\mathbb{N})$: both x and y belong to $\mathcal{P}^{\mathbf{m+n}}(\mathbb{N})$, and so $\{x, y\} \in \mathcal{P}^{\mathbf{m+n+1}}(\mathbb{N})$: the structure satisfies the axiom of elementary sets. If $A \in \mathcal{P}^{\mathbf{i}}(\mathbb{N})$, then $\mathcal{P}(A) \in \mathcal{P}^{\mathbf{i+1}}(\mathbb{N})$. If $A \in \mathcal{P}^{\mathbf{i}}(\mathbb{N})$ (for $i > 0$), then $\bigcup A \in \mathcal{P}^{\mathbf{i-1}}(\mathbb{N})$: the restriction to positive i is no real restriction because $\mathbb{N} \subseteq \mathcal{P}(\mathbb{N})$. Infinity obviously holds since $\mathbb{N}$ belongs to the structure. If it is supposed that Choice holds in the whole universe it certainly holds in this structure, as a choice set for a partition in $\mathcal{P}^{\mathbf{i+1}}(\mathbb{N})$ will belong to $\mathcal{P}^{\mathbf{i}}(\mathbb{N})$

Notice the similarity between the role of iterated power sets of the natural numbers in our description of this structure and types in the theory of the previous section. The only difference is that the analogues of types here are cumulative.

4.1.1 Exercises

1. We define x^+ as $x \cup \{x\}$. We use the modern form of the Axiom of Infinity: there is a set which contains $\emptyset$ and is closed under $x \mapsto x^+$. We implement 0 as $\emptyset$, and if the natural number n is implemented as the set x, $n+1$ is implemented as x^+.

   We define $\mathbb{N}$ as the intersection of all sets which contain 0 and are closed under successor. Explain how we can show that this set exists using the axioms of infinity and separation.

   Show that the axioms of Peano arithmetic are satisfied in this implementation of $\mathbb{N}$. Proofs of axioms 1,2,3,5 should be very straightforward.

   Axiom 4 requires you to show that $x \cup \{x\} = y \cup \{y\}$ implies $x = y$ for all $x, y \in \mathbb{N}$. Show this using the axioms of Zermelo set theory (*without*

Foundation).

Hints: how do you prove *anything* about natural numbers? You can begin as an exercise by proving that for no natural number n is $n \in n$ true, by induction of course. This is similar to the fact about natural numbers you need to prove to establish Axiom 4. I will give more explicit hints if you visit me with work in progress.

2. Write a proof in Zermelo set theory with the modern form of the Axiom of Infinity (and without Foundation) that no natural number is an element of itself. This will of course be an induction proof using the definitions $0 = \emptyset; n+1 = n^+ = n \cup \{n\}$. Intense attention to "obvious" detail is needed at this level. Hint: it will be useful (and easy) to prove first (by induction of course) that all natural numbers are transitive.

   Even more of a hint: the induction step looks like this. Suppose $n \notin n$. Our goal is to show $n+1 = n \cup \{n\}$ is not an element of itself. Suppose otherwise for the sake of a contradiction. We suppose that is that $n+1 \in n+1 = n \cup \{n\}$. So either $n+1 \in n$ (something bad happens...) or $n+1 = n$ (something bad happens...).

4.2 The intended interpretation of Zermelo set theory in set pictures; the Axiom of Rank; transitive closures and Foundation

Our intention in this section is to show how Zermelo set theory can be interpreted in subsets of the set Z of set pictures with the relation E standing in for membership, and to observe that when Zermelo set theory is implemented in this way certain additional axioms hold which make the system easier to work with.

Any sentence of the language of untyped set theory can be translated into a sentence of our type theory by replacing each occurrence of $\in$ with the relation E and bounding each quantifier in the set Z (all in some fixed type). In fact, instead of bounding it in Z, we bound it in $\mathbb{E}_\lambda$, where $\lambda \geq \omega \cdot 2$ is a limit ordinal. We assume that each rank below rank λ is complete, so we are assuming at least the existence of $\beth_\omega$.

We claim that (under the assumption that all types below λ are complete), the translations of the axioms of Zermelo set theory into the language of type

theory are true, so we have a way to understand untyped set theory in terms of our type theory.

Extensionality: Sets with the same elements are the same. It appears that Zermelo may allow atoms (non-sets) in his original formulation, but we will assume here that all objects are sets.

Verification of Extensionality: This follows from the fact that E is a membership diagram, and so an extensional relation (and the fact that E end extends the restriction of E to any $\mathbb{E}_\lambda$; the preimage of any element of the field of the restriction under the restriction is the same as its preimage under E itself, so extensionality of E implies extensionality of the restriction.

Elementary Sets: The empty set $\emptyset$ exists. For any objects x and y, $\{x\}$ and $\{x, y\}$ are sets.

Verification of Elementary Sets: The equivalence class of the empty set diagram belongs to $\mathbb{E}_\lambda$ and has empty preimage under E, so satisfies the translation of the properties of the empty set. Let $a, b \in \mathbb{E}_\lambda$. $a \in \mathbb{E}_\alpha$ and $b \in \mathbb{E}_\beta$ for some $\alpha, \beta < \lambda$. Because λ is limit, $\max(\alpha, \beta) + 1 < \lambda$, and since $\mathbb{E}_{\max(\alpha,\beta)}$ is a complete rank, $\{a, b\}$ has an E-code in $\mathbb{E}_{\max(\alpha,\beta)+1} \subseteq \mathbb{E}_\lambda$

Separation: For any property $\phi[x]$ and set A, the set $\{x \in A \mid \phi[x]\}$ exists.

Verification of Separation: Any $A \in \mathbb{E}_\lambda$ belongs to some rank $\mathbb{E}_{\alpha+1}$ for $\alpha < \lambda$ (every element of Z first appears in a successor rank). The formula $\phi[x]$ translates to a formula $\Phi[x]$ in the language of type theory. The set $\{x \in \mathbb{E}_\alpha \mid x\, E\, A \wedge \Phi[x]\}$ exists by comprehension in type theory and has an E-code in $\mathbb{E}_{\alpha+1}$ because $\mathbb{E}_\alpha$ is a complete rank.

Power Set: For any set A, the set $\{B \mid B \subseteq A\}$ exists. The definition of $A \subseteq B$ is the usual one.

Verification of Power Set: Any $A \in \mathbb{E}_\lambda$ belongs to some rank $\mathbb{E}_{\alpha+1}$ for $\alpha < \lambda$ (every element of Z first appears in a successor rank). $\alpha + 1$ and $\alpha + 2$ are also less than λ because λ is limit. The translation of $B \subseteq A$ asserts that the E-preimage of B is a subset of the E-preimage of A. Each B whose E-preimage is a subset of the E-preimage of A

also belongs to $\mathbb{E}_{\alpha+1}$ (because each element of its E-preimage belongs to $\mathbb{E}_\alpha$ and $\mathbb{E}_\alpha$ is complete), so the set of all such B has an E-code in $\mathbb{E}_{\alpha+2}$, because $\mathbb{E}_{\alpha+1}$ is complete.

Union: For any set A, the set $\bigcup A = \{x \mid (\exists y \in A.x \in y)\}$ exists.

Verification of Union: Any $A \in \mathbb{E}_\lambda$ belongs to some rank $\mathbb{E}_{\alpha+1}$ for $\alpha < \lambda$ (every element of Z first appears in a successor rank). The translation of $(\exists y \in A.x \in y)$ into the language of type theory asserts that x is in the $E|E$-preimage of A, which is a subset of $\mathbb{E}_\alpha$, so has an E-code in $\mathbb{E}_{\alpha+1}$, because $\mathbb{E}_\alpha$ is complete.

Infinity: There is a set I such that $\emptyset \in I$ and $(\forall x.x \in I \to \{x\} \in I)$.

Verification of Infinity: Define a relation on the ordinals $\leq \omega$ by $x\,R\,y \leftrightarrow y = x + 1 \vee y = \omega$. The isomorphism type of this relation is the implementation of I.

Choice: Any pairwise disjoint collection of nonempty sets has a choice set.

Verification of Choice: The translation of the property "P is a pairwise disjoint collection of nonempty sets" into the language of type theory is "P is an element of $\mathbb{E}_\lambda$ such that the E-preimages of the elements of its E-preimage are nonempty and disjoint". $P \in \mathbb{E}_\lambda$ belongs to some rank $\mathbb{E}_{\alpha+1}$ for $\alpha < \lambda$. Each element of the E-preimage of an element of the E-preimage of P belongs to $\mathbb{E}_\alpha$. By the Axiom of Choice in type theory, the pairwise disjoint collection of nonempty E-preimages of the elements of the E-preimage of P has a choice set, which is a subset of $\mathbb{E}_\alpha$, so has an E-code because $\mathbb{E}_\alpha$ is a complete rank.

Furthermore, the translation of the axioms of Zermelo set theory into the theory of all set pictures expressed with type-free set picture variables are true, with a qualification, for essentially the same reasons given above. The qualification is that Separation will only work for formulas in which every quantifier is bounded in a set, because we cannot translate sentences which do not have this property from the language of type-free set picture variables back into the language of type theory. The version of Zermelo set theory with this restriction on Separation is called "bounded Zermelo set theory" or "Mac Lane set theory", the latter because Saunders Mac Lane has advocated it as a foundational system. Notice that the translation of Mac Lane set theory

into the type-free theory of set pictures does *not* require the assumption that $\beth_\omega$ exists: the only axiom that requires that λ be limit in the development above is Power Set, and the verification of the translation of Power Set in the theory of all set pictures is given at the end of the previous section (basically, one can introduce the "power set" of any particular "set" one mentions by working in a higher type).

We state an additional axiom which holds in both the implementations of Zermelo set theory given here, but which fails to hold in some eccentric models of Zermelo set theory. This axiom expresses the idea that every element of $\mathbb{E}_\lambda$ belongs to some rank $\mathbb{E}_\alpha$.

Observation: The Kuratowski pair $\{\{x\}, \{x, y\}\}$ of two sets x and y is easily seen to be a set, and the proof that this is a pair goes much as in type theory. We can then define relations (and in particular well-orderings) just as we did in type theory.

Definition: A *subhierarchy* is a set H which is well-ordered by inclusion and in which each successor in the inclusion order on H is the power set of its predecessor and each limit in the inclusion order on H is the union of all its predecessors in that order. A *rank* is a set which belongs to some subhierarchy.

Theorem: Of any two distinct subhierarchies, one is an initial segment of the other in the inclusion order. So all ranks are well-ordered by inclusion.

Axiom of Rank: Every set is a subset of some rank.

Verification of the Axiom of Rank: Each $A \in \mathbb{E}_\lambda$ belongs to some $\mathbb{E}_\alpha$, $\alpha < \lambda$. Each $\mathbb{E}_\alpha$ has an E-code, which we will call V_α, because it is a complete rank. For any β, $\{V_\alpha \mid \alpha < \beta\}$ has an E-code, which we will call H_α, because it is a subset of $\mathbb{E}_{\beta+1}$. It is straightforward to verify that H_α satisfies the stated properties for a subhierarchy (translated into the language of type theory), whence we have the translation of "V_α is a rank", and "$A \subseteq V_\alpha$", so the translation of "A is a subset of some rank" holds.

The Axiom of Rank has many useful consequences. We give two of them here.

Definition: We say that a set A is *transitive* iff $(\forall x \in A.(\forall y \in x.y \in A))$. It is worth noting that a set is transitive (in our interpretation in type theory) iff any set diagram belonging to the set picture implementing A is a transitive relation.

Theorem: Every set is included in a transitive set.

Proof: It is straightforward to prove by transfinite induction along the inclusion order that all ranks are transitive. By the Axiom of Rank every set is included in a rank.

Definition: For any set A, we define r_A as the minimal rank in the inclusion order including A as a subset. We define $\mathtt{TC}(A)$, the *transitive closure* of A, as $\{x \in r_A \mid$ every transitive set including A includes $x\}$. This exists by Separation and is the minimal transitive set in the inclusion order which includes A as a subset. $\mathtt{TC}(\{A\})$, which also contains A as an element, will sometimes be of more interest.

Observation: That sets have transitive closures is not provable in Zermelo set theory as originally formulated. The usual proof in *ZFC* requires the very powerful Axiom of Replacement. This is deceptive, as Zermelo set theory with the Axiom of Rank is not essentially stronger than Zermelo set theory (it is possible to interpret the latter in the former), while the Axiom of Replacement makes Zermelo set theory far stronger.

Theorem (the Axiom of Foundation): Every set A has an element x such that x is disjoint from A.

Proof: Let r be the minimal rank in the inclusion order which includes an element of A as an element, and let $x \in r \cap A$. Each element of x belongs to a rank properly included in r, so x is disjoint from A.

The Axiom of Foundation is frequently (anachronistically) adjoined to Zermelo set theory as an additional axiom.

We observed above that the modern form of the Axiom of Infinity and the original form do not imply each other in the presence of the other axioms. They do imply each other in the presence of the Axiom of Rank. For the Axiom of Rank, combined with the existence of any infinite set, implies that there is a minimal infinite rank V_ω in the inclusion order, and both the Zermelo natural numbers and the von Neumann natural numbers are

definable subsets of V_ω (since all of the elements of either are clearly of finite rank). It is also amusing to note that the Axioms of Pairing and Union can be omitted in the presence of the Axiom of Rank. $\{a, b\}$ can be derived using Separation as a subset of the power set of $r_a \cup r_b$ (this binary set union exists because it is actually one of the ranks r_a and r_b), and $\bigcup A$ can be derived using Separation as a subset of $\mathtt{TC}(A)$.

4.3 Developing mathematics in Zermelo set theory

In this section we develop basic mathematical constructions in Zermelo set theory.

We begin with a very basic

Theorem: $(\forall A.(\exists x.x \notin A))$

Proof: This theorem follows from Separation alone. Consider $R_A = \{x \in A \mid x \notin x\}$. Suppose $R_A \in A$. It follows that $R_A \in R_A \leftrightarrow R_A \notin R_A$.

Since it follows from Foundation that $x \notin x$ for any x, it further follows from Zermelo set theory with the Axiom of Rank that $R_A = A$ for all A.

It is a fundamental characteristic of Zermelo set theory (and of all stronger theories) that there are no very big sets (such as the universe V). Many mistake this for a fundamental characteristic of set theory.

We want to implement relations and functions. Here it is very convenient to work with the Kuratowski pair.

Definition: $\langle x, y\rangle = \{\{x\}, \{x, y\}\}$

Theorem: For any sets x and y, $\langle x, y\rangle$ is a set.

This theorem is not enough by itself to ensure that we can use the Kuratowski pair to get an adequate theory of relations.

Definition: $A \cup B = \bigcup\{A, B\}$

Theorem: $A \cup B$ exists for any sets A and B (this is clear from the form of the definition). $A \cup B = \{x \mid x \in A \vee x \in B\}$. Notice that the latter definition, which we used as the primary definition in type theory, is *not* guaranteed to define a set by Separation.

Definition: $A \cap B = \{x \in A \mid x \in B\}$; $A - B = \{x \in A \mid x \notin B\}$. If A is a set and $B \in A$, $\bigcap A = \{x \in B \mid (\forall a \in A. x \in A)\}$.

Theorem: For any sets A and B, $A \cap B$ and $A - B$ exist. This is obvious from the forms of the definitions. If A is nonempty, $\bigcap A$ exists and the definition of the set does not depend on the choice of the element B.

Definition: $A \times B = \{x \in \mathcal{P}^{\mathbf{2}}(A \cup B) \mid (\exists a \in A. (\exists b \in B. x = \{\{a\}, \{a, b\}\}))\}$.

Theorem: $A \times B$ exists for all sets A and B. $A \times B = \{\langle a, b\rangle \mid a \in A \wedge b \in B\}$. The existence of $A \times B$ is obvious from the form of the definition. The trick is to notice that any pair $\langle a, b\rangle = \{\{a\}, \{a, b\}\}$ with $a \in A$ and $b \in B$ actually belongs to $\mathcal{P}^{\mathbf{2}}(A \cup B)$, because $\{a\}$ and $\{a, b\}$ both belong to $\mathcal{P}(A \cup B)$.

Definition: A *relation* is a set of ordered pairs. We define $x\, R\, y$ as $\langle x, y\rangle \in R$.

Observation: Just as in type theory, not every logical relation is a set relation. For example, the logical relation of equality is not implemented as a set, because $Q = \{\langle x, y\rangle \mid x = y\}$ would have the unfortunate property $\bigcup^{\mathbf{2}} Q = V$, the universal set, which we know does not exist. For similar reasons, membership and inclusion are not set relations.

Definition: For any relation R, we define $\mathtt{fld}(R)$ as $\bigcup^{\mathbf{2}} R$, $\mathtt{dom}(R)$ as

$$\{x \in \mathtt{fld(R)} \mid (\exists y. \langle x, y\rangle \in R)\},$$

and $\mathtt{rng}(R)$ as

$$\{y \in \mathtt{fld(R)} \mid (\exists x. \langle x, y\rangle \in R)\}.$$

Theorem: The field, domain, and range of a relation R are sets. This is evident from the forms of the definitions. That they are the intended sets is evident from the fact that if $\langle x, y\rangle \in R$ then $x, y \in \bigcup^{\mathbf{2}} R$. Moreover, $\mathtt{fld}(R) = \mathtt{dom}(R) \bigcup \mathtt{rng}(R)$ and $R \subseteq \mathtt{dom}(R) \times \mathtt{rng}(R) \subseteq \mathtt{fld}(R)^2$.

Once we have verified that we have an adequate foundation for the theory of relations, we can import definitions and concepts wholesale from type theory, always subject to the limitation that we cannot construct very large collections. For example we cannot define cardinals, ordinals, or general

isomorphism types as equivalence classes under equinumerousness or isomorphism, because equinumerousness, isomorphism, and most of their equivalence classes are not sets.

In Zermelo set theory as originally formulated, there is no uniform solution to this problem. However, in Zermelo set theory with the Axiom of Rank we have a formal device to remedy this lack, known as "Scott's trick".

Definition: For any formula $\phi[x]$, define $r_{\phi[x]}$ as the minimal rank r such that $(\exists x \in r.\phi[x])$, or as the empty set if there is no such rank r. Define $\{x : \phi\}$ as $\{x \in r_{\phi[x]} \mid \phi[x]\}$. $\{x : \phi[x]\}$ is obviously a set for all formulas $\phi[x]$.

Definition: $A \sim B$ iff there is a bijection from A onto B, as in type theory. $|A|$, the *Scott cardinal* of A, is defined as $\{B : B \sim A\}$. For any relations R and S, we say that $R \approx S$ iff there is a bijection f from $\mathtt{fld}(R)$ onto $\mathtt{fld}(S)$ such that $x\,R\,y \leftrightarrow f(x)\,S\,f(y)$. We define the *Scott isomorphism type* of R as $\{S : S \approx R\}$. Scott isomorphism types of well-orderings are called *Scott order types* (of the well-orderings: $\mathtt{ot}(W)$ is the Scott order type of a well-ordering W) or *Scott ordinal numbers* (as a class).

Now that we have defined ordinals we can define the notation V_α.

Definition: For any subhierarchy h, introduce the nonce notation $\leq_h$ for the inclusion order restricted to h. If α is an ordinal we define V_α as the rank A (if there is one) such that $\mathtt{ot}((\leq_h)_A) = \alpha$ for any $\leq_h$ with A in its field. It is straightforward to show that $\mathtt{ot}((\leq_h)_A)$ is the same ordinal for any $\leq_h$ with A in its field, and that A is uniquely determined by α.

In Zermelo set theory with the Rank Axiom we can prove that every set belongs to some V_α but we cannot prove the existence of $V_{\omega \cdot 2}$. But we do have the ability to define order types for every well-ordering and cardinals for every set using the Scott definitions.

These are not the definitions of cardinal and ordinal number which are usually used in *ZFC*. We give those definitions (due to von Neumann) but they have the limitations that they do not necessary work in Zermelo set theory without Replacement (not all well-orderings can be shown to have

order types, nor can all sets be shown to have cardinals) and the von Neumann definition of cardinal depends essentially on the Axiom of Choice, as the Scott definition does not.

Definition: A *(von Neumann) ordinal number* is a transitive set which is strictly well-ordered by the membership relation.

Observation: In our implementation of Zermelo set theory in set pictures, a von Neumann ordinal number α is implemented by the isomorphism type of strict well-orderings of type $\alpha + 1$ (except for 0, which is implemented by the order type of the usual empty order). In $\mathbb{E}_{T^2(\lambda)}$, only the ordinals less than λ are implemented in this way. If $\lambda = \omega \cdot 2$, this definition is not useful: the only infinite well-orderings with order types are of the form $\omega + n$, but there are much longer order types that are realized (such as ω_1). A hypothesis adequate to make this definition useful is "$\beth_{T^2(\lambda)}$ exists for each ordinal λ" in the ambient type theory. The Axiom of Replacement of *ZFC* makes this definition usable (and is much stronger).

Definition: The *(von Neumann) order type* of a well-ordering W is the von Neumann ordinal α such that the union of the restrictions of the membership and equality relations on α^2 is isomorphic to W.

Definition: The *(von Neumann) cardinality* of a set A is the smallest von Neumann ordinal which is the order type of a well-ordering of A.

We make a general claim here that mathematical results can be imported from type theory to untyped set theory. It is useful to give a uniform account of how such a general claim can be justified (which also makes it clear exactly what is claimed).

Just as we can translate the language of Zermelo set theory into the language of type theory in a way which makes the axioms true, so we can translate the language of type theory into the language of untyped set theory in a way which makes the axioms true – and so makes all the theorems true.

Let ϕ be a formula of the language of type theory mentioning n types. Let $X_0, X_1, \ldots X_{n-1}$ be a sequence of sets such that $\mathcal{P}(X_i) \subseteq X_{i+1}$ for each appropriate index i. The translation $(\phi)_X$ is defined as follows: each quantifier over type i is restricted to X_i. Each formula $x \in y$, where x is of type i and y is of type $i + 1$ is translated as $x \in y \wedge y \in \mathcal{P}(X_i)$ (elements of

$X_{i+1} - \mathcal{P}(X_i)$ are interpreted as urelements); formulas of the form $x = y$ are interpreted as $x = y$. Such a translation is also feasible if there is an infinite sequence with the same properties, but it is not a theorem of Zermelo set theory that there are such sequences. A specific version which we will write $[\phi]_X$ has $X_i = \mathcal{P}^{\mathbf{i}}(X)$ for a fixed set X: a nice feature of this version is that we can generate as many terms of the sequence as we need in a uniform way. It is straightforward to verify that as long as X_0 is infinite the translations of all axioms of type theory into the language of untyped set theory are true. It can further be noted that expressions T representing sets in the language of type theory will also have translations $(T)_X$ where X is a sequence or $[T]_X$ where X is a set.

This makes a wide class of mathematical assertions readily portable from type theory to set theory. For example, all of our assertions about cardinal and ordinal arithmetic have readily determined analogues in untyped set theory.

4.4 Digression: Interpreting typed set theory as Mac Lane set theory

Mac Lane set theory (untyped set theory with the boundedness restriction on the Axiom of Separation) can be interpreted in typed set theory with strong extensionality, using our entire universe of typed objects. We begin by postulating an operator J which is injective ($J(x) = J(y) \to x = y$) and sends type 0 objects to type 1 objects. An example of such an operator is the singleton operator ι. Any such J can be thought of as implemented by a function $\iota``V^{\mathbf{0}} \to V^{\mathbf{1}}$.

We now indicate how to extend the J operator to all types. If J is defined for type n objects we define $J(x^{\mathbf{n+1}})$ as $\{J(y^{\mathbf{n}}) \mid y^{\mathbf{n}} \in x^{\mathbf{n+1}}\}$. Briefly, $J(x) = J``x$. It is easy to see that J is injective on every type: we have $J(x) = J(y) \leftrightarrow x = y$, no matter what the common type of x and y. By the definition of J at successive types, we further have $J(x) \in J(y) \leftrightarrow x \in y$, no matter what the successive types of x and y.

In our interpretation of untyped set theory, we identify every object x of whatever type with each of its iterated images $J^n(x)$ under the J operator: in this way each type n is seen to be embedded in type $n + 1$. If x is of type m and y is of type n, we have $x = y$ in the interpretation iff $J^n(x) = J^m(y)$ (note that both of these terms are of the same type $m + n$) and we have

$x \in y$ in the interpretation iff $J^n(x) \in J^{m+1}(y)$ (in which the terms have successive types $m+n$ and $m+n+1$). Notice that if x and y are of the same type n, $J^n(x) = J^n(y) \leftrightarrow x = y$, and if x and y are of successive types n and $n+1$, $J^{n+1}(x) \in J^{n+1}(y) \leftrightarrow x \in y$: where equality and membership make sense in type theory, they coincide with equality and membership in the typed theory.

If x, y, z have types m, n, p, and we have $x = y$ and $y = z$, we have $J^n(x) = J^m(y)$ and $J^p(y) = J^n(z)$. Further applications of J to both sides of these formulas show that transitivity of equality works: $J^{n+p}(x) = J^{m+p}(y)$ and $J^{m+p}(y) = J^{m+n}(z)$ are implied by the previous equations and imply $J^{n+p}(x) = J^{m+n}(z)$, which in turn by injectivity of J implies $J^p(x) = J^m(z)$ which is the interpretation of $x = z$. Reflexivity and symmetry of equality present no difficulties. The substitution property of equality requires some technical detail for its verification which we do not (NOTE: yet) give here.

We verify that some of the axioms of Mac Lane set theory hold in this interpretation.

We discuss the Axiom of Extensionality. Suppose that x is of type n and y is of type $n+k$. If $k = 0$ and x and y have the same elements, then $x = y$ by the axiom of extensionality of type theory. Otherwise, if for all z of type m we have $z \in x$ iff $z \in y$ in the interpreted theory, this means we have $J^n(z) \in J^m(x)$ iff $J^{n+k}(z) \in J^m(y)$, and further $J^{n+k}(z) \in J^{m+k}(x)$, whence $J^m(y) = J^{m+k}(x)$, whence $J^n(y) = J^{n+k}(x)$, whence $x = y$ in the interpretation, which is what is wanted.

Now we discuss the Axiom of Bounded Separation. We want to show the existence of $\{x \in A \mid \phi[x]\}$ in the untyped theory, where ϕ is a formula in membership and equality (it should not mention the predicate of typehood, which does not translate to anything in the language of type theory, though of course it may mention specific types) we suppose that every quantifier in $\phi[x]$ is restricted to a set. Assign referents to each free variable appearing in $\{x \in A \mid \phi[x]\}$, then assign each bound variable the type one lower than that assigned to the set to which it is restricted (A in the case of x, the bound on the quantifier in the case of quantified sets; if the bound is type 0, make the variable type 0 as well), then apply our interpretation of the untyped language in the typed language (adding applications of J to variables in such a way as to make everything well-typed). For example, $\{x \in A \mid x \notin x\}$ would become $\{x \in A \mid x \notin J(x)\}$, with x being assigned type one lower than that assigned to A (unless A was assigned a referent of type 0, in which case we would have $\{x \in J(A) \mid x \notin J(x)\}$). The resulting set abstract exists

in our typed theory and has the right extension in the interpretation. If there were unbounded quantifiers in $\phi[x]$, there would be no way to interpret them in terms of our typed theory, which does not allow any way to quantify over objects of all types.

(NOTE: more axioms to be supplied. Rank will not necessarily hold here; the form of infinity which holds depends on the exact form of J. This development is more *ad hoc* and more closely related to the original form(s) of Zermelo set theory).

Something like this interpretation can also be carried out in the version of type theory with weak extensionality. We detail the modifications of the construction.

The operation J must be defined at atoms in each positive type. J is defined on type 0 as an injective operation raising type by 1, as above. If J is defined on type n objects, we define it on type $n+1$ *sets* as before: $J(x^{\mathbf{n+1}}) = J``(x^n)$. There are no more than $T|V^{\mathbf{n+1}}|$ elementwise images under J in type $n+2$: since $T|V^{\mathbf{n+1}}| < |\mathcal{P}(V^{\mathbf{n+1}})|$ by Cantor's theorem, we can choose as many distinct further elements of $\mathcal{P}(V^{n+1})$, i.e., *sets* in V^{n+2}, as we need as images of the type $n+1$ atoms under J. The result $x \in y \leftrightarrow J(x) \in J(y)$ now holds only if y is a set, and for this reason we modify the interpretation of $x \in y$ in the untyped theory (where x and y have types m and n respectively in type theory) to $J^{n+1}(x) \in J^{m+1}(y)$; if x were of type 0 and y were an urelement of whatever type the original interpretation $J^n(x) \in J^m(y)$ would not work correctly.

4.5 The von Neumann definitions of ordinal and cardinal number

We introduced the perhaps mysterious traditional definition of *ordinal number* due to von Neumann above:

Definition: An *ordinal number* is a transitive set A which is strictly well-ordered by membership (i.e., the restriction of the membership relation to $A \times A$ is the strict partial order corresponding to a well-ordering). Or "a transitive set of transitive sets none of which are self-membered".

Observation: This seems to be equivalent to "x is an ordinal iff x is a transitive set, no element of x is self-membered, and x is well-ordered

by inclusion". This has the merit that our preferred definition of well-ordering is used. Let x be an ordinal by this definition. For each $y \in x$, and each $z \in y$, we have $z \in x$ because x is transitive, so we have either $z \subset y$ or $y \subseteq z$. But $y \subseteq z$ is impossible because this would imply $z \in z$.

Definition: For any ordinal α, use $\in_\alpha$ to represent $\in \cap \alpha^2$ (which we know is a strict well-ordering).

Theorem: For any ordinal α and any $\beta \in \alpha$, β is an ordinal, $\beta = \mathtt{seg}_{\in_\alpha}(\beta)$ and $\in_\beta = (\in_\alpha)_\beta$.

Proof: $\delta \in \gamma \in \beta \rightarrow \delta \in_\alpha \gamma \in_\alpha \beta$ ($\gamma, \delta \in \alpha$ because α is transitive) and this implies $\delta \in_\alpha \beta$ and so $\delta \in \beta$ because $\in_\alpha$ is a partial order. Thus $\beta \in \alpha$ is transitive. $\in \cap \beta^2$ is a strict well-ordering because it is a suborder of $\in \cap \alpha^2$. Further, it is evident that $\in_\beta = (\in_\alpha)_\beta$: the order on β is the segment restriction of the order on α determined by β, because β is identical to the segment in the order on α determined by β: $\beta = \{\gamma \in \alpha \mid \gamma \in \beta\}$ (this uses transitivity of α) $= \{\gamma \mid \gamma \in_\alpha \beta\}$.

Theorem: For any two ordinal numbers α and β, exactly one of the following is true: $\alpha = \beta$, $\alpha \in \beta \wedge \alpha \subseteq \beta$, $\beta \in \alpha \wedge \beta \subseteq \alpha$. Any set of ordinal numbers is thus linearly ordered by $\subseteq$: moreover, this linear order is a well-ordering.

Proof: By a basic theorem on well-orderings proved above, we know that there is either an isomorphism from $\in_\alpha$ to $\in_\beta$, an isomorphism from $\in_\alpha$ to some $(\in_\beta)_\gamma = \in_\gamma$ for some $\gamma \in \beta$ or an isomorphism from $\in_\beta$ to some $(\in_\alpha)_\gamma = \in_\gamma$ for some $\gamma \in \alpha$. It is then clearly sufficient to show that for any ordinals α and β, if $\in_\alpha \approx \in_\beta$, then $\alpha = \beta$. Suppose for the sake of a contradiction that $f : \alpha \rightarrow \beta$ is an isomorphism from $\in_\alpha$ to $\in_\beta$ and that there is some $\gamma \in \alpha$ such that $f(\gamma) \neq \gamma$. There is then a $\in_\alpha$-least such γ. We have γ as the $\in_\alpha$-least element of α such that $f(\gamma) \neq \gamma$. The objects which are $\in_\beta f(\gamma)$ are exactly those $f(\delta)$ such that $\delta \in_\alpha \gamma$ (this is just because f is an isomorphism). We can read $\in_\alpha$ and $\in_\beta$ simply as membership, and we remind ourselves that for any $\delta < \gamma$ $f(\delta) = \delta$, and thus we see that $\gamma = f(\gamma)$ because they have the same members, which is a contradiction.

That $\alpha \in \beta \to \alpha \subseteq \beta$ expresses the fact that ordinals are transitive sets. $\subseteq$ is a partial order on sets and what we have shown so far indicates that it is a linear order on ordinals. To see that it is a well-ordering, we need to show that any nonempty set A of ordinals has a $\subseteq$-least element: since A is nonempty, we can choose $\alpha \in A$; either there is some $\beta \in \alpha$ which is an element of A or there is not. If there is none, then α is the $\subseteq$- (and $\in$-) least element of A; otherwise the $\subseteq -$ (and $\in$-) least element of α which belongs to A will be the $\subseteq -$ (and $\in$-) least element of A: there is such an element because $\in$ is a strict well-ordering of α and so $\subseteq$ is a well-ordering of α.

Definition: For any well-ordering $\leq$, we define $\mathtt{ot}(\leq)$ as the ordinal α (if any) such that the well-ordering of α by $\subseteq$ is isomorphic to $\leq$.

Definition: For any set A, we define $|A|$ as the minimal ordinal α in the inclusion order such that $A \sim \alpha$.

Note that in the usual set theory we identify a cardinal number with its initial ordinal: these are not the same object in type theory, though of course they are closely related. This is another of those differences between possible implementations of mathematical concepts in set theory that one should watch out for (in the Scott implementation of cardinals and ordinals in Zermelo set theory, a cardinal is not identified with its initial ordinal). The fact that though we identify these concepts formally in *ZFC* we do not actually think of them as having the same mathematical content is witnessed by the fact that we use different notations for $\mathbb{N}$ (the set of natural numbers), ω (the first infinite ordinal) and $\aleph_0$ (the first infinite cardinal) although these are all implemented as exactly the same object! Note that in type theory they are all different.

The axioms we have given so far do not ensure that all well-orderings have order types or that all sets have cardinalities: we have already described intepretations of the axioms of Zermelo set theory in type theory in which these statements do not hold.

The axioms of Zermelo set theory (as given here) do ensure that each finite well-ordering has an order type, and each finite set has a cardinality. So we have already provided a full implementation of the natural numbers using the von Neumann definitions. Further, one can deduce from the modern version of the Infinity axiom of Zermelo set theory (and Separation) that the first

infinite ordinal ω exists as a set, which by the conventions we have given is to be identified with both the cardinal $\aleph_0$ and the set $\mathbb{N}$ of all natural numbers.

It is important to notice that just as there can be no set V of all sets in Zermelo set theory, there can be no set `Ord` of all ordinals (so transfinite induction and recursion must be stated in property-based or restricted forms in this theory). For the ordinals are strictly well-ordered by membership in an obvious external sense: if there were a set Ω which contained all ordinals, it would be an ordinal, so we would have $\Omega \in \Omega$, and this is impossible again by the definition of ordinal. This is a version of the Burali-Forti paradox, another of the classical paradoxes of set theory.

4.5.1 Exercises

1. The Scott definition of a natural number n is that it is the collection of all sets of size n and rank as low as possible. Remember the rank of a set A is the first ordinal α such that A is a subset of V_α. Write down as many Scott natural numbers as explicit sets as you can stand to. Work out the sizes of the next few (how many elements do they have? – go up to 20 or so?) All you need for this is an understanding of what $V_0, V_1, V_2 \ldots$ (the finite ranks) are, and some familiar combinatorics. You might also want to see what you can say about the Scott natural number 60000 versus the Scott natural 70000. There is a dramatic difference (smiley).

2. The Axiom of Foundation asserts that for any nonempty set x there is a set $y \in x$ such that $x \cap y = \emptyset$.

   One way of understanding this is that this axiom says that if we look at $\in \cap x^2$ (the membership relation on x^2 that it must have a "minimal" element – "minimal" is in scare quotes because membership is not an order relation. A "minimal" element y will have empty preimage under the membership relation restricted to x – that is, it wil have no elements in common with x.

   Use the Axiom of Foundation (along with the other axioms of course) to prove the following:

   (a) There is no set x such that $x \in x$.

   (b) There is no sequence s such that for all $n \in \mathbb{N}$ we have $s_{n+1} \in s_n$.

The strategy to follow is this: in each part, identify a set which would have no "minimal" element in the membership relation.

4.6 The Axiom of Replacement and *ZFC*

We introduce the missing assumption of the usual set theory which makes it possible to prove that the von Neumann definitions of ordinal and cardinal number are total.

class function notation: If we have a formula $\phi[x, y]$ such that for every x there is at most one y such that $\phi[x, y]$, we introduce notation $y = F_\phi(x)$ for the unique y associated with a given x. Notice that F_ϕ is not understood to be a set here.

Axiom of Replacement: If we have a formula $\phi[x, y]$ such that for every x there is at most one y such that $\phi[x, y]$, and define $F_\phi(x)$ as above, then for every set A, the set $\{F_\phi(x) \mid x \in A\}$ exists.

The Axiom of Replacement can be used then to justify the recursive definition of the V_α's above. What the axiom of replacement says, essentially, is that any collection we can show to be the same size as or smaller than a set is in fact a set.

Theorem: $\mathtt{ot}(\leq)$ exists for every well-ordering $\leq$.

Proof: Let $\leq$ be a well-ordering such that $\mathtt{ot}(\leq)$ does not exist. If there are x such that $\mathtt{ot}((\leq)_x)$ does not exist, define $\leq_0$ as $(\leq)_x$ for the smallest such x; otherwise define $\leq_0$ as $\leq$ itself. In either case $\leq_0$ is a well-ordering which has no order type with the property that all of its initial segments have order types. We now define a formula $\phi[x, \alpha]$ which says "α is the order type of $(\leq_0)_x$" (the tricky bit is showing that we can say this). Notice that once we do this we are done: $\{F_\phi(x) \mid x \in \mathtt{fld}(\leq_0)\}$ will be the first von Neumann ordinal after all the order types of segment restrictions of $\leq_0$, which will be the order type of $\leq_0$ contrary to assumption.

$\phi[x, \alpha]$ is defined as "if f is a (set) function with domain an initial segment of $\leq_0$ containing x and having the property $f(y) = \{f(z) \mid z \leq_0 y\}$ for each y in its domain, then $f(x) = \alpha$". It is straightforward

to prove that exactly one such function f exists for each initial segment of $\leq_0$ (its extendability at limits in $\leq_0$ uses Replacement).

We have already seen that provision of this formula leads to a contradiction to our original assumption.

Corollary: The von Neumann cardinal $|A|$ exists for every set A.

Proof: There is a well-ordering of A, whose order type is an ordinal with the same cardinality of A. Either this is the smallest ordinal (in the inclusion order) with this property, in which case it is $|A|$ itself, or it has elements which have this property, among which there must be a smallest, which is $|A|$.

Theorem: V_α exists for each α.

Proof: Consider the smallest ordinal λ for which V_λ does not exist (it is obviously a limit ordinal if it exists).

Find a formula $\phi[\alpha, A]$ which says "$A = V_\alpha$" and we are done, because we can then define the set $\{F_\phi(\alpha) \mid \alpha \in \lambda\}$, and the union of this set will be V_λ contrary to assumption.

The formula $\phi[\alpha, A]$ says "there is a function f whose domain is an ordinal β such that $\alpha \in \beta$, and $f(0) = \emptyset$, $f(\gamma + 1) = \mathcal{P}(f(\gamma))$ if $\gamma + 1 \in \beta$, and $f(\mu) = \bigcup\{f(\gamma) \mid \gamma \in \mu\}$ for each limit ordinal $\mu \in \beta$, and $f(\alpha) = x$". The fact that there is a unique such function f for each $\beta < \lambda$ is readily shown: Replacement is used to show extendability of f at limit ordinals.

Zermelo set theory augmented with the Axiom of Replacement is known as *ZFC* (Zermelo-Fraenkel set theory with Choice). [The Rank Axiom is not needed because it can be proved from the other axioms of Zermelo set theory and the Axiom of Replacement.] This is the system of set theory which is most commonly used.

Although the Axiom of Replacement is sufficient to make the von Neumann definitions of cardinality and order type succeed, it is certainly not necessary. A weaker axiom with the same effect (already noted above in a type-theoretic form) is

Axiom of Beth Numbers: For every Scott ordinal α, $\beth_\alpha$ exists.

We define things in terms of Scott ordinals because we do not wish to presume that the von Neumann ordinal α exists; that is what we are trying to prove. A set of size $\beth_\alpha$ must be included in a rank V_β with $\beta \geq \alpha$, and the von Neumann ordinal α will be present in $V_{\beta+1}$. Notice that the Axiom of Rank plays an essential role in this argument: existence of large $\beth$ numbers in the original Zermelo set theory does not have any effect on existence of von Neumann ordinals.

Another axiom which works is the stronger

Axiom of Beth Fixed Points: For every cardinal κ, there is a cardinal $\lambda > \kappa$ such that $\beth_\lambda = \lambda$.

4.7 Translation between Type Theory and Set Theory

We discuss how to transfer mathematical concepts and theorems from type theory to set theory.

We have already seen that any formula of the language of type theory can be translated to a formula $[\phi]_X$ (where X is an infinite set) which asserts that ϕ holds in a model of type theory in which X is type 0, $\mathcal{P}(X)$ is type 1, and in general $\mathcal{P}^{\mathbf{n}}(X)$ is type n. $[\phi]_X$ is obtained by rereading membership and equality as the relations of the untyped theory and restricting each type n variable to $\mathcal{P}^{\mathbf{n}}(X)$. For each axiom ϕ of type theory (in each of its explicitly typed versions), it is straightforward to show that $[\phi]_X$ is a theorem of Zermelo set theory. So for any theorem ϕ of type theory we have $[\phi]_X$ a theorem of Zermelo set theory, and in fact we also have "for all infinite sets X, $[\phi]_X$" a theorem of Zermelo set theory.

Every object t we can define in the language of type theory has analogues t_X for each infinite set X. This presents an obvious problem (a stronger version of the ambiguity of type theory which our avoidance of type indices partially obscures). All our definitions of specific objects, with a few exceptions such as $\emptyset$, refer to different objects depending on the choice of the parameter X. For example the number $3^{\mathbf{n+2}}$ is implemented as $[\mathcal{P}^{\mathbf{n}}(X)]^3$, the set of all subsets of $\mathcal{P}^{\mathbf{n}}(X)$ with exactly three elements. Just which set this is varies with the choice of X (and n).

A possible conceptual problem with the theory of functions can be dispelled: in type theory, we can prove easily that the functions from a set A to a set B defined using Kuratowski pairs correspond precisely to those defined

using Quine pairs (they are at different types but this ceases to be so inconvenient when we are translating to untyped set theory). So the question of which sets are the same size and which relations are isomorphic is settled in the same way no matter which pair definition one uses.

Nonetheless, the theory of cardinals and ordinals can be stated in untyped set theory as the theory of specific objects. Here we suppose that we use von Neumann ordinals as the implementation of ordinal numbers, and von Neumann initial ordinals as the implementation of cardinals. A sentence $(|A| = \kappa)_X$ asserts that A belongs to a certain cardinal κ_X. This translates to an assertion $|A| = \kappa$ in the language of untyped set theory, now not meaning $A \in \kappa_X$ but $A \sim \kappa$, where κ is the first von Neumann ordinal which is equinumerous with an element (and so with all elements) of κ_X. Further, it is important to note that for any cardinal $(\kappa^{\mathbf{n}})_X$ the von Neumann initial ordinal associated with it will be the same as the von Neumann initial ordinal associated with $(T(\kappa)^{\mathbf{n+1}})_X$: this gives a concrete meaning to our erstwhile intuitive feeling that κ and $T(\kappa)$ are in fact the same cardinal. Very similar considerations apply to order types $(\alpha)_X$ and corresponding von Neumann ordinals α (and we get the analogous result that the ordinals $(\alpha^{\mathbf{n}})_X$ and $(T(\alpha)^{\mathbf{n+1}})_X$ correspond to the same von Neumann ordinal α). Further, nothing but technical points would differ if we used the Scott cardinals and ordinals here instead of the von Neumann cardinals and ordinals. Since we have a translation of ordinals and cardinals of type theory to ordinals and cardinals of the untyped set theory, we can translate the operations of addition and multiplication from type theory to untyped set theory directly. It might seem that we cannot translate cardinal exponentiation so directly, but here we observe that though $(\kappa^{\lambda})_X$ is not always defined, it is always the case that $(T(\kappa)^{T(\lambda)})_X$ is defined (and will be $T(\kappa^{\lambda})_X$ if the latter is defined); since the T operation is now understood to be the identity, we see that cardinal exponentiation is now a total operation. The way in which the definitions of cardinals and ordinals are transferred from type theory to set theory ensures that theorems of cardinal and ordinal arithmetic transfer as well. Notice that Cantor's Theorem now takes the form $\kappa < 2^{\kappa}$: there is no largest cardinal (from which it follows that there can be no universal set, as certainly $|V| \geq 2^{|V|}$ would hold; the argument from the untyped form of Cantor's theorem and the naive supposition that there is a universal set to a contradiction is called *Cantor's paradox*).

Although we have just defined operations of cardinal and ordinal arithmetic in terms of the interpreted type theory with X as type 0, it is perfectly

possible to state definitions of these operations which do not depend on the notation $[\phi]_X$. The recursive definitions of operations of ordinal arithmetic are inherited directly by untyped set theory from type theory. The definitions of $|A| + |B|$ as $|(A \times \{0\} \cup B \times \{1\}|$, $|A| \cdot |B|$ as $|A \times B|$, and $|A|^{|B|}$ as $|A^B|$ work perfectly well in untyped set theory (always remembering that the set theoretical meaning of $|A|$, though not its mathematical function, is quite different). But the correspondence between the arithmetic of interpreted type theory and the arithmetic of untyped set theory is important in seeing that theorems can be relied upon to transfer fairly directly from type theory to set theory.

Results we have given above imply that certain statements which can be shown to be true in the version of Zermelo set theory interpreted in our type theory with strong extensionality are inconsistent with *ZFC*. We showed above that $\beth_\alpha$ does not exist for some α in these models (to be precise, if the cardinality of the set corresponding to type 0 is $\aleph_\beta$, we can prove that $\beth_{\beta+\omega}$ does not exist in that model (whereas in *ZFC* we have $|V_{\omega+\alpha}| = \beth_\alpha$ for each ordinal α, so all $\beth_\alpha$'s must exist)) However, there are models of Zermelo set theory obtained from models of type theory with weak extensionality in which *ZFC* holds. This might seem not be possible since there is a sequence of sets $V^{\mathbf{i}}$ (the sets corresponding to the types) such that any cardinal is less than some $|V^{\mathbf{i}}|$ (since it is the cardinal of a set of some type): by Replacement it might seem that the countable sequence of $V^{\mathbf{i}}$'s would be a set (because it is the same size as the set of natural numbers), so its union would be a set, which would have cardinality larger than any $|V^{\mathbf{i}}|$. But this argument does not work, because there is no formula defining the sequence of $V^{\mathbf{i}}$'s (as there is in the models based on type theory with strong extensionality, where $V^{\mathbf{i+1}} = \mathcal{P}(V^{\mathbf{i}})$). Moreover, we will apply simple model theory below to show that for any model of *ZFC* there is a model obtainable from a model of type theory with weak extensionality in which the same statements of the language of set theory are true [that is a very convoluted sentence, I know].

The serious difference in power between untyped set theory and typed set theory has to do with the ability to quantify over the entire universe. This is just a difference in what we can *say* if we use Bounded Separation, but if we adopt the full axiom of Separation we can define sets in terms of global facts about the universe. This is best indicated using an example.

Theorem: For each natural number n, there is a unique sequence s of sets with domain the set of natural numbers $\leq n$ such that $s_0 = \mathbb{N}$ and for

each $i < n$, $s_{i+1} = \mathcal{P}^n(\mathbb{N})$.

Proof: Prove this by mathematical induction. The set of natural numbers n for which there is such a sequence s clearly includes 0 ($s = \langle 0, \mathbb{N}\rangle$) and if it includes k will also include $k+1$ (if s works for k, $s \cup \langle k+1, \mathcal{P}(s_k)\rangle$ works for $k+1$).

Discussion: In type theory with base type countable, sets interpreting these sequences do not all exist in any one type, so no assertion of type theory can even express the fact that they all exist. This statement is of course very badly typed, but a similar assertion would be the statement that there is a sequence of cardinals such that $s_0 = \aleph_0$ and $s_{i+1} = 2^{s_i}$ for each i, and this would present the same problem: in type theory with base type countable, the sequence $\beth_0, \beth_1 \beth_2, \ldots$ is not entirely present in any one type. The mere statement of the theorem cannot be expressed in type theory because the quantifier over sequences s is not bounded in a set (and for this same reason this theorem cannot be proved using Bounded Separation: the subset of the natural numbers which needs to be shown to be inductive cannot be shown to exist).

4.7.1 Exercises

1. This proof will use Replacement.

   In the usual axiom set it is rather more involved than it seems it ought to be to show that every set is a subset of a transitive set (this is easily shown in cumulative type theory or in Zermelo set theory with the rank axiom, but the usual formulation of Zermelo set theory or *ZFC* has the Foundation axiom, which is weaker.

   I give an outline of a proof which you need to complete (there are models in the notes for the proof).

   Let X be a set. We want to prove that there is a transitive set which contains X. The idea is to prove that the collection of sets $\{\bigcup^{\mathbf{n}}(X) \mid n \in \mathbb{N}\}$ exists. Then you can show that the union of this set is transitive and contains X as a subset.

   Fill in the details. To prove the existence of $\{\bigcup^{\mathbf{n}}(X) \mid n \in \mathbb{N}\}$ by Replacement you need a formula $\phi[x, n]$ which says "$x = \bigcup^{\mathbf{n}} X$". As I said, there are models for this in the notes.

Why does it follow immediately from "X is a subset of a transitive set" that X is an element of some transitive set as well?

2. This question is intended to address the question of just how weird a model of *ZFC* without the Axiom of Rank can be.

   Work in *ZFC*. Define a set A as *bounded* iff its transitive closure contains finitely many von Neumann natural numbers. We refer to the first von Neumann natural not in the transitive closure of A as the bound of A. Verify the following points:

   (a) If a and b are bounded, $\{a, b\}$ is bounded.

   (b) If A is bounded, $\mathcal{P}(A)$ is bounded (but the bound might go up by one – do you see why?), and $\bigcup A$ is bounded (with the same bound). You should also show that the bounds of the sets $\mathcal{P}^k(A)$ eventually increase with k.

   (c) The set of Zermelo natural numbers is bounded.

   (d) If the bound of A is n, the set $\mathcal{P}^{>n}(A)$ of all subsets of A with more than n elements is also bounded with bound n (note that this can be iterated).

   (e) Apply the points above to argue that the collection of all bounded sets in the universe of *ZFC* is a model of Zermelo set theory in which the set of von Neumann natural numbers does not exist, in which $\{\mathcal{P}^n(X) \mid n \in \mathbb{N}\}$ does not exist for any X, and in which there are sets of every cardinality which exists in the universe of *ZFC*.

5 Logic

5.1 Formalization of Syntax and Substitution

In this section we discuss the representation of bits of syntax (formulas and terms) by mathematical objects. We will thereafter identify the syntactical objects of our language with these mathematical objects.

An obvious way to do this would be to represent ASCII characters by natural numbers, then represent character strings as functions from finite initial

segments of $\mathbb{N}$ to ASCII characters. But the definition of formal operations on syntax with this definition would be inconvenient.

Our representation will be typically ambiguous, as with all our representations of mathematical objects in type theory: syntactical objects will exist in all types above a certain minimum type (which we really will not care about determining). Though we work in type theory it should be clear how the same construction could be done in Zermelo set theory.

We initially give a recursive definition of notation taken from logic and set theory as mathematical objects.

We begin with variables. The triple $\langle 0, m, n\rangle$ will represent a bound variable $x_n^{\mathbf{m}}$ and the triple $\langle 1, m, n\rangle$ will represent a free variable (or "constant") $a_n^{\mathbf{m}}$ for natural numbers m, n. The reasons why we want bound and free variables will become evident later. That is, we define 'x_n^m' as $\langle 0, m, n\rangle$ and 'a_n^m' as $\langle 1, m, n\rangle$.

The triple $\langle 2, n, t\rangle$, where t is a term, will represent the sentence $P_n(t)$ (P_n being a unary predicate). The quadruple $\langle 3, n, t, u\rangle$ will represent the sentence $t\, R_n\, u$ (R_n being a binary predicate (logical relation)). We read $\langle 3, 0, t, u\rangle$ as $t \in u$ and $\langle 3, 1, t, u\rangle$ as $t = u$. That is, we define '$P_n(t)$' as $\langle 2, n, \text{'}t\text{'}\rangle$ and '$t\, R_n\, u$' as $\langle 3, n, \text{'}t\text{'}, \text{'}u\text{'}\rangle$.

The triple $\langle 4, n, t\rangle$ (t being a term) stands for $F_n(t)$ (F_n being a function symbol). The quadruple $\langle 5, n, t, u\rangle$ (t and u being terms) stands for $t\, O_n\, u$, O_n being a binary function (operation) symbol. That is, '$F_n(t)$' is defined as $\langle 4, n, \text{'}t\text{'}\rangle$ and '$t\, O_n\, u$' is defined as $\langle 5, n, \text{'}t\text{'}, \text{'}u\text{'}\rangle$.

Note that all predicate and function symbols are typically ambiguous (can be used with arguments of many types). Binary relation symbols are assumed to be type level and functions are assumed to have one or both inputs and their output all of the same type.

The triple $\langle 6, n, t\rangle$ represents $\iota^{\mathbf{n}}(t)$ and the triple $\langle 7, n, t\rangle$ represents $\bigcup^{\mathbf{n}} t$. Note that we can now represent $t \in u$ as $\{t\} \subseteq u$. That is, '$\iota^{\mathbf{n}}(t)$' is defined as $\langle 6, n, \text{'}t\text{'}\rangle$ and '$\bigcup^{\mathbf{n}} t$' is defined as $\langle 7, n, \text{'}t\text{'}\rangle$.

The quadruple $\langle 8, n, v, \phi\rangle$ (where ϕ is a formula and v is a variable) is read $(Q_n v.\phi)$, where Q_n is a quantifier. We reserve Q_0 as $\exists$ and Q_1 as $\forall$. That is, '$(Q_n v.\phi)$' is defined as $\langle 8, n, \text{'}v\text{'}, \text{'}\phi\text{'}\rangle$.

The quadruple $\langle 9, n, v, \phi\rangle$ (where ϕ is a formula and v is a variable) represents a term $(B_n v.\phi)$ constructed by binding on a formula. We read $\langle 9, 0, v, \phi\rangle$ as $(\epsilon v.\phi)$, the Hilbert symbol. Note that $\{v \mid \phi\}$ can be read as $(\epsilon A.(\forall v.\{v\} \subseteq A \leftrightarrow \phi))$. That is, $(\epsilon v.\phi)$ (in particular) is defined as $\langle 9, 0, \text{'}v\text{'}, \text{'}\phi\text{'}\rangle$.

The pair $\langle 10, \phi \rangle$ represents $\neg\phi$. The triple $\langle 11, \phi, \psi \rangle$ represents $\phi \vee \psi$. That is, '$\neg\phi$' is defined as $\langle 10, \text{`}\phi\text{'} \rangle$ and '$\phi \vee \psi$' is defined as $\langle 11, \text{`}\phi\text{'}, \text{`}\psi\text{'} \rangle$.

The above is not precisely mathematical as it relates mathematical objects to pieces of notation. We proceed to develop a thoroughly mathematical account of syntax and semantics using this informal account as motivation. For readability, we will allow ourselves to use quoted terms and formulas much of the time.

Definition: This is a nonce definition. A syntactical pair of sets is a pair of sets $\langle T, F \rangle$ with the following properties, motivated by the idea that T is an approximation to the set of terms and F is an approximation to the set of formulas.

1. For any natural numbers m, n, $\langle 0, m, n \rangle$ and $\langle 1, m, n \rangle$ belong to T. Objects $\langle 0, m, n \rangle$ are called bound variables.
2. For any natural number n and any $t \in T$, $\langle 2, n, t \rangle \in F$.
3. For any natural number n and $t, u \in T$, $\langle 3, n, t, u \rangle \in F$.
4. For any natural number n and $t \in T$, $\langle 4, n, t \rangle$, $\langle 6, n, t \rangle$, $\langle 7, n, t \rangle \in T$
5. For any natural number n and $t, u \in T$, $\langle 5, n, t, u \rangle \in T$.
6. For any natural number n, , bound variable v, $\phi \in F$, and $t, u \in T$, $\langle 8, n, v, \phi \rangle \in F$, and $\langle 9, n, v, \phi \rangle \in T$.
7. For any $\phi, \psi \in F$, $\langle 10, \phi \rangle \in F$ and $\langle 11, \phi, \psi \rangle \in F$.

Definition: A *formal term set* is any set which is the first projection T of a syntactical pair $\langle T, F \rangle$. A *formal proposition set* is any set which is the second projection F of a syntactical pair $\langle T, F \rangle$. A *formal term* is an object which belongs to all formal term sets. A *formal proposition* is an object which belongs to all formal proposition sets.

Theorem: If $\mathcal{T}$ is the set of all formal terms and $\mathcal{F}$ is the set of all formal propositions, then $\langle \mathcal{T}, \mathcal{F} \rangle$ is a syntactical pair of sets.

The two sets $\mathcal{T}$ and $\mathcal{F}$ are defined by mutual recursion. It is natural to prove theorems about formal terms and propositions using structural induction. We will write formal terms and propositions using ordinary typography, and in fact to the best of our ability forget the intricacies of numerals and pairing that underly the formal definition (particularly since the details are

largely arbitrary and could be changed wholesale without affecting the subsequent development).

Terms have type, and considerations of type determine that some terms are ill-formed. $x_n^{\mathbf{m}}$ and $a_n^{\mathbf{m}}$ have type m. $F(t)$ has the same type as t. $t\,O\,u$ has type m iff t and u have the same type m and is ill-typed otherwise. $(\epsilon x_n^{\mathbf{m}}.\phi)$ (this is the Hilbert symbol) has type m. A formula $t\,R\,u$ will only be considered well-formed if t and u have the same type. If t has type n, $\iota^{\mathbf{k}}(t)$ has type $n+k$ and $\bigcup^{\mathbf{k}}(t)$ has type $n-k$ if $n \geq k$ and is considered ill-formed otherwise. These clauses are enough to determine the typing (and well-formedness) of all terms and formulas by recursion.

Now we give the formal definition of substitution. We define $u[t/x_i]$ (the result of replacing x_i with t in the term u) and $\phi[t/x_i]$ (the result of replacing x_i with t in the formula ϕ) at the same time. Here we leave off the type index: the type requirement is that t and x_i have the same type.

1. $x_j[t/x_i]$ is defined as t if $i = j$ and as x_j otherwise.
2. $a_j[t/x_i]$ is defined as a_j.
3. $F(u)[t/x_i]$ is defined as $F(u[t/x_i])$.
4. $(u\,O\,v)[t/x_i]$ is defined as $u[t/x_i]\,O\,v[t/x_i]$.
5. $(Bx_j.\phi)[t/x_i]$ is defined as $(Bx_k.\phi[x_k/x_j][t/x_i])$, where x_k is the first variable not found in $(Bx_j.\phi)[t/x_i]$. The full definition of $(Bu.\phi)[t/x_i]$ where u may be a complex term is quite difficult. The only form that B takes in our development is the Hilbert symbol ϵ.
6. $P(u)[t/x_i]$ is defined as $P(u[t/x_i])$.
7. $(u\,R\,v)[t/x_i]$ is defined as $u[t/x_i]\,R\,v[t/x_i]$.
8. $(Qx_j.\phi)[t/x_i]$ is defined as $(Qx_k.\phi[x_k/x_j][t/x_i])$, where x_k is the first variable not occurring in $(Qx_j.\phi)[t/x_i]$. Defining $(Qu.\phi)[t/x_i]$ for general terms u would be difficult.
9. $(\neg\phi)[t/x_i]$ is defined as $\neg\phi[t/x_i]$ and $(\phi \vee \psi)[t/x_i]$ is defined as $\phi[t/x_i] \vee \psi[t/x_i]$.

To justify that this definition works takes a little thought. The notion of length of a term or formula can be defined by a natural recursion (we do not give the mind-numbing details here). Then observe that the substitution of t for x_i in any given formula P is may be defined in terms of other substitutions supposed already defined, but these are always substitutions into strictly shorter formulas.

Our formulation of syntax differs from usual formulations in defining a single universal formal language, which is specifically adapted to the needs of type theory, though it can also be used for single-sorted first order theories. The adaptation to first-order theories is straightforward: simply do not use variables of type other than zero or the singleton or union operations. The language would need to be extended for more complicated multi-sorted theories (more complicated type theories): we will not discuss this. The language could be extended with n-ary predicate and function symbols for $n > 2$, of course. It can obviously be cut down by specifying limited collections of constants, unary and binary predicate symbols, and unary and binary function symbols.

5.1.1 Exercises

1. Using the definitions of formal syntax above, write out the mathematical object coding the formula

$$(\forall x_1^3.(\exists x_2^3.x_1^3\, R_2\, x_2^3)).$$

2. What is the term or formula coded by

$$\langle 8, 1, \langle 0, 0, 1\rangle, \langle 3, 2, \langle 0, 0, 1\rangle, \langle 1, 0, 1\rangle\rangle\rangle?$$

5.2 Formalization of Reference and Satisfaction

In this section we define the notions of *meaning* and *truth*. That is, given an interpretation of the nonlogical symbols of our language, we show how to formally define the referent of each term and the truth value of each formula, mod assignments of values to all variables.

We first need to set the stage. A domain of objects is needed to support our interpretation. In fact, we supply a sequence D_n of domains, one for each $n \in \mathbb{N}$, with D_n intended to be the collection of type n objects. Note that all the sets D_n are actually of the same type in the sense of our working type

theory. If we restrict our language to the first-order as indicated above, we only need a single domain D.

We associate a value $a_i^n \in D_n$ with each constant $a_i^{\mathbf{n}}$. With each unary predicate P_i we associate a set $P_i^n \subseteq D_n$ for each n (because our language is typically ambiguous we need an interpretation of each predicate over each type). With each binary relation symbol R_i we associate a set $R_i^n \subseteq D_n \times D_n$ for each n. Similarly each unary function symbol F_i is associated with functions $F_i^n : D_n \to D_n$, and each binary operation symbol O_i with functions $O_n : D_n^2 \to D_n$. An injective map $\iota_{n+1} : D_n \to D_{n+1}$ is provided for each n, and a map $\bigcup_n : D_{n+1} \to D_n$ with the property $\bigcup_n(\iota_{n+1}(x)) = x$ for each $x \in D^n$. (The existence of these latter maps imposes requirements on the sequence of sets D_n: the sets in the sequence must be of increasing size). To support the Hilbert symbol we provide a function H_n from nonempty subsets of D_n for each n: $H_n(A) \in 1$ for all A; if $A \subseteq D_n$, then $H_n(A) \subseteq A$ if $A \neq \emptyset$; $H_n(\emptyset)$ is defined and belongs to ι"D_n but is otherwise unspecified.

A *structure* for our formal language is determined by a map D sending a possibly proper initial segment of the natural numbers to domains D_n, "singleton" and "union" maps $\iota^{\mathbf{n+1}} : D^n \to D^{n+1}$ and $\bigcup^{\mathbf{n}} : D^{n+1} \to D^n$ as above, modified choice functions H_n as above (if the Hilbert symbol is to be used), and some partial functions implementing constants, predicates and functions as indicated above: where m, n are natural numbers, $A(m, n)$ will be the element a_n^m of D_n used as the referent of $a_n^{\mathbf{m}}$, $P(m, n)$ will be the subset P_n^m of D^m intended to be the extension of the predicate P_n in type m, $R(m, n)$ will be the subset R_n^m of D_m^2 intended to be the extension of the logical relation R_n, $F(m, n)$ is the element F_n^m of $D_m^{D_m}$ representing the action of the function symbol F_n in type m, and $O(m, n)$ is the element O_n^m of $D_m^{D_m^2}$ representing the action of the operation symbol O_n in type m. The length of the domain sequence and the domain of the partial function determine the subset of our universal language which is used in the obvious way.

The binding constructions used in the discussion which follows are limited. The only term construction binding propositions we provide is the Hilbert symbol $(\epsilon x.\phi[x])$ which may be read "an x such that $\phi[x]$ if there is one (chosen in an unspecified manner if there are more than one) or a default object if there is no such x". All definable term binding constructions (including the set builder notation) can be defined in terms of the Hilbert operator. The only quantifiers we provide are the usual ones (which can in

fact also be defined in terms of the Hilbert operator!). It is not difficult to extend the discussion to general binders, but it would further complicate already very elaborate recursive definitions.

A possibly partial function E on variables such that $E(x_i^{\mathbf{n}}) \in D_n$ for each variable $x_i^{\mathbf{n}}$ in the domain of E is called an *environment*. If E is an environment we define $E[d/x_i^{\mathbf{n}}]$ as the environment which sends $x_i^{\mathbf{n}}$ to d and agrees with E everywhere else (this may be an extension of E if E is not defined at $x_i^{\mathbf{n}}$).

We will now recursively define functions $\mathcal{R}$ and $\mathcal{V}$ (named with "reference" and "valuation" in mind). These functions take two arguments, an environment and a term. They are partial functions: they are sometimes undefined. Strictly speaking, these functions are defined relative to a structure and would be written $\mathcal{R}_{\mathcal{S}}$ and $\mathcal{V}_{\mathcal{S}}$ if we wanted to explicitly specify a structure S we were working with. We use the informal notation a_i^n for $S(a_i^{\mathbf{n}})$, P_i^n for $S(n, P_i)$, and so forth. The domains of these functions are restricted to the language appropriate to the structure (and further restricted depending on the extent to which E is partial).

We define $v(\phi)$ as 1 if ϕ is true and 0 if ϕ is false.

1. $\mathcal{R}(E, x_i^{\mathbf{n}}) = E(x_i^{\mathbf{n}})$ (if this is defined).
2. $\mathcal{R}(E, a_i^{\mathbf{n}}) = A(n, i)$.
3. $\mathcal{R}(E, F_i(t))$ is defined as $F(n, i)(\mathcal{R}(E, t))$, where n is the type of t (as long as $\mathcal{R}(E, t)$ is defined).
4. $\mathcal{R}(E, \iota^{\mathbf{k}}(t))$ is defined as $\iota_{n+k}(\mathcal{R}(E, \iota^{\mathbf{k-1}}(t)))$, where n is the type of t, as long as the embedded reference is defined. $\iota^{\mathbf{0}}(t)$ is to be read as just t.
5. $\mathcal{R}(E, \bigcup^{\mathbf{k}}(t))$ is defined as $\bigcup_{n-k}(\mathcal{R}(E, \bigcup^{\mathbf{k-1}}(t)))$, where n is the type of t, as long as the embedded reference is defined. $\bigcup^{\mathbf{0}}(t)$ is to be read as just t.
6. $\mathcal{R}(u\, O_i\, v)$ is defined as $O(n, i)(\mathcal{R}(E, u), \mathcal{R}(E, v))$ just in case $\mathcal{R}(E, u)$ and $\mathcal{R}(E, v)$ are defined and u and v have the same type n.
7. $\mathcal{R}(E, (\epsilon x_i^{\mathbf{n}}.\phi))$ is defined as the sole element of $H_n(\{d \in D_n \mid \mathcal{V}(E[d/x_i^{\mathbf{n}}], \phi) = 1\})$ if the valuation is defined.

8. $\mathcal{V}(E, P(u))$ is defined as $v(\mathcal{R}(E,u) \in P(n,i))$, where n is the type of u, as long as the embedded reference is defined.

9. $\mathcal{V}(E, (u\, R_i\, v)$ is defined as $v(\langle \mathcal{R}(E,u), \mathcal{R}(E,v)\rangle \in R(n,i))$, as long as the embedded references are defined and u and v have the same type n.

10. $\mathcal{V}(E, (Qx^{\mathbf{n}}_j.\phi))$ is defined as $v((Qd \in D_n.\mathcal{V}(E[d/x^{\mathbf{n}}_i], \phi) = 1))$, where Q is either $\exists$ or $\forall$, as long as the embedded valuation is defined.

11. $\mathcal{V}(E, \neg\phi)$ is defined as $v(\neg(\mathcal{V}(E,\phi) = 1))$, and $\mathcal{V}(E, \phi \vee \psi)$ is defined as $v(\mathcal{V}(E,\phi) = 1 \vee \mathcal{V}(E,\psi) = 1)$, as long as the embedded valuations are defined.

Notice as with substitution that the reference and valuation functions are defined recursively. Reference and valuation for a particular term or formula may appeal to reference or valuation for another formula or term, but always a strictly shorter one.

Although our language is restricted for convenience in framing these definitions, the full language of type theory is supported with suitable definitions. If equality and subset relations are primitive, we define $x \in y$ as $\iota(x) \subseteq y$, $\phi \to \psi$ as $\neg\phi \vee \psi$, $\phi \wedge \psi$ as $\neg(\neg\phi \vee \neg\psi)$, $\phi \leftrightarrow \psi$ as $\phi \to \psi \wedge \psi \to \phi$, and $\{x \mid \phi\}$ as $(\epsilon A.(\forall x.x \in A \leftrightarrow \phi))$.

5.2.1 Exercises

1. Use the definitions of reference and satisfaction to evaluate the following expressions, if $D_0 = \{1,2,3\}$ and the following information about the environment and interpretation is given. Notice that we really do not need to worry about types in this example.

   $A(0,1) = 3$ (that is, the intended referent of a^0_1 is 3).

   $P(0,1) = \{1,2\}$

   $R(0,1) = \{\langle 1,1\rangle, \langle 2,2\rangle, \langle 3,3\rangle\}$ (the equality relation).

   $E(x^0_n) = 1$ for all n (the environment E assigns every type 0 variable the value 1).

   Show the reasoning behind your evaluation in detail. The intended evaluations are quite obvious: the point is to show that the nasty defi-

nitions in the notes actually get us there, so detail must be seen. This is an exercise in step by step unpacking of definitions.

(a) $\mathcal{R}(E, a_1^0)$

(b) $\mathcal{R}(E, x_5^0)$

(c) $\mathcal{V}(E, P_1(a_1^0))$

(d) $\mathcal{V}(E, P_1(x_1^0))$

(e) $\mathcal{V}(E, x_2^0\, R_1\, a_1^0)$

(f) $\mathcal{V}(E, x_2^0\, R_1\, x_5^0)$

(g) $\mathcal{V}(E, (\exists x_2^0 . x_2^0\, P_1\, a_1^0))$

5.3 Formal Propositional Sequent Calculus

We introduce sequent notation.

Definition: A *sequent* is an ordered pair $\langle \Gamma, \Delta \rangle$ of finite sets of formulas. We write sequents $\Gamma \vdash \Delta$. The set $\{A\}$ (where A is a formula) is simply written A in a sequent; the set $\Gamma \cup \{A\}$ is written Γ, A; notation for the empty set is omitted.

Definition: A sequent $\Gamma \vdash \Delta$ is *valid* iff every interpretation under which $\mathcal{V}$ is defined for all elements of Γ and Δ [we will presume this condition for all interpretations and sequents hereinafter] and under which $\mathcal{V}$"$\Gamma \subseteq \{1\}$ has $1 \in \mathcal{V}$"Δ (every interpretation which makes *all* statements in Γ true makes *some* statement in Δ true).

Lemma: $\Gamma, A \vdash \Delta, A$ is a valid sequent for any formula A and sets Γ and Δ.

Lemma: $\Gamma, \neg A \vdash \Delta$ is a valid sequent iff $\Gamma \vdash A, \Delta$ is a valid sequent.

Lemma: $\Gamma \vdash \neg A, \Delta$ is a valid sequent iff $\Gamma, A \vdash \Delta$ is a valid sequent.

Lemma: $\Gamma, A \vee B \vdash \Delta$ is a valid sequent iff both $\Gamma, A \vdash \Delta$ and $\Gamma, B \vdash \Delta$ are valid sequents. Note that this is a formalized version of the strategy of proof by cases.

Lemma: $\Gamma \vdash A \vee B, \Delta$ is a valid sequent iff $\Gamma \vdash A, B, \Delta$ is a valid sequent.

We introduce a weaker notion of valuation appropriate when we are considering propositional logic only.

Definition: A *propositional valuation* is a partial function $\mathcal{V}$ which sends each formula in its domain to either 0 or 1, and which sends any formula $\neg\phi$ to $1 - \mathcal{V}(\phi)$ and any formula $\phi \vee \psi$ to $\mathcal{V}(\phi) + \mathcal{V}(\psi) - \mathcal{V}(\phi) \cdot \mathcal{V}(\psi)$ (in each case iff the valuations of subformulas are defined).

Observation: All valuations in the sense of the previous section are propositional valuations, but not vice versa.

Definition: A propositionally valid sequent is one in which any propositional valuation which is defined on all formulas involved and sends all formulas on the left to 1 sends some formula on the right to 1. Note that all propositionally valid sequents will be valid, but not vice versa (a formula which is not propositionally valid may be valid for other logical reasons).

Observation: All the Lemmas above remain true when "valid" is replaced with "propositionally valid".

Theorem: If a sequent ϕ is propositionally valid, applications of the rules above will inevitably show this. If a sequent ϕ is not propositionally valid, applications of the rules above will inevitably reduce the sequent to a form from which a valuation witnessing its invalidity can be extracted.

Proof: Any application of the rules above converts a sequent with n disjunctions and negations in it to one or two sequents with $n-1$ disjunctions and negations each. So sufficiently many applications of the rules will convert any sequent into a collection of sequents in which all formulas are atomic (or quantified), but in any event do not have accessible disjunctions or negations. If each of these sequents has a formula in common between its left and right sets, the sequent is valid. If one of these sequents does not have a formula in common between its left and right sides, a valuation assigning 1 to each formula on the left and 0 to each formula on the right witnesses the fact that the original formula is (propositionally) invalid. The total number of steps will be no more than $1 + 2 + 4 + \ldots + 2^n = 2^{n+1} - 1$ (which means that proofs of complex sequents may be impractically large!), because if we start with a

sequent with n connectives and organize our work into steps in which we apply a single rule to each sequent, at step k we will obtain no more than 2^k formulas of length $n - k$.

So we have given a complete formal account of propositional logic.

It is worth noting that a form of the rules above can be given in which all sequents have the empty set or a singleton set on the right. Many readers will be comfortable with many premisses but only a single intended conclusion (the case of the empty set represents the goal of a contradiction). This can be done purely mechanically: apply the rules in the forms given above, then, if there is more than one formula on the right, convert all but one of them to their negations and move them to the left. In the case of the negation rule, move the original conclusion to the left; in the case of the right rule for disjunction, move the second disjunct to the left. The theorem still holds.

The given rules can be used to derive rules for the other propositional connectives. These resemble the proof strategies that we have developed in the section on Proof, with the notable exception that the left rule for implication seems different (although it does support the modus ponens and modus tollens strategies we expect). The resemblance of the sequent rules to our proof strategies is clearer in the single-conclusion forms (though the left rule for implication remains eccentric).

We can present sequent proofs as mathematical objects.

Definition: An axiom (a sequent with nonempty intersection between the left and right side) is a proof of its own validity.

If the validity of sequent A follows from the validity of sequent B by an application of a sequent rule, and C is a proof of B, then $\langle A, C\rangle$ is a proof of A.

If the validity of sequent A follows from the validity of sequents B abd C by an application of a sequent rule, and D is a proof of A and E is a proof of C, then $\langle A, \langle D, E\rangle\rangle$ is a proof of A.

Being an instance of one of the sequent rules is mathematically definable, so the notion of being a sequent proof is mathematically definable (the class of sequent proofs is the smallest class with the closure conditions just described).

Note that the addition of more sequent rules will cause only minor adjustments to this definition.

A sequent is *provable* if there is a proof of it. A sentence ϕ is provable iff the sequent $\vdash \phi$ is provable.

We give the propositional sequent rules in a useful format. In each entry, the validity of the sequent below the line is equivalent to all the sequents above the line being valid.

$$\Gamma, A \vdash A, \Delta$$

$$\frac{\Gamma \vdash A, \Delta}{\Gamma, \neg A \vdash \Delta}$$

$$\frac{\Gamma, A, B \vdash \Delta}{\Gamma, A \wedge B \vdash \Delta}$$

$$\frac{\Gamma, A \vdash \Delta \qquad \Gamma, B \vdash \Delta}{\Gamma, A \vee B \vdash \Delta}$$

$$\frac{\Gamma \vdash A, \Delta \qquad \Gamma, B \vdash \Delta}{\Gamma, A \to B \vdash \Delta}$$

$$\frac{\Gamma, A \to B, B \to A \vdash \Delta}{\Gamma, A \leftrightarrow B \vdash \Delta}$$

$$\frac{\Gamma, A \vdash \Delta}{\Gamma \vdash \neg A, \Delta}$$

$$\frac{\Gamma \vdash A, B, \Delta}{\Gamma \vdash A \vee B, \Delta}$$

$$\frac{\Gamma \vdash A, \Delta \qquad \Gamma \vdash B, \Delta}{\Gamma \vdash A \wedge B, \Delta}$$

$$\frac{\Gamma, A \vdash B, \Delta}{\Gamma \vdash A \to B, \Delta}$$

$$\frac{\Gamma, A \vdash B, \Delta \qquad \Gamma, B \vdash A, \Delta}{\Gamma \vdash A \leftrightarrow B, \Delta}$$

5.4 Formal First-Order Sequent Calculus: the Completeness, Compactness and Löwenheim-Skolem Theorems

For first-order reasoning, we need to introduce sequent rules for quantification.

Lemma (Cut Rule): $\Gamma \vdash \Delta$ is valid iff $\Gamma, A \vdash \Delta$ and $\Gamma \vdash A, \Delta$ are both valid.

This may seem like a purely propositional rule, though we did not need it in the previous section. As we will see in a later subsection, we do not need it here either, but it is very convenient.

We give the sequent rules for quantifiers (and the Hilbert symbol).

Lemma: $\Gamma, (\exists x.\phi[x]) \vdash \Delta$ is valid iff $\Gamma, \phi[a] \vdash \Delta$ is valid, where a is a constant which does not appear in the first sequent.

Lemma: $\Gamma \vdash (\forall x.\phi[x]), \Delta$ is valid iff $\Gamma \vdash \phi[a], \Delta$ is valid, where a is a constant which does not appear in the first sequent.

Lemma: $\Gamma \vdash (\exists x.\phi[x]), \Delta$ is valid iff $\Gamma \vdash \phi[t], (\exists x.\phi[x]), \Delta$ is valid, where t is any term.

Lemma: $\Gamma, (\forall x.\phi[x]) \vdash \Delta$ is valid iff $\Gamma, (\forall x.\phi[x]), \phi[t] \vdash \Delta$ is valid, where t is any term.

The sequent rules for equality are the following.

Lemma: Any sequent of the form $\Gamma \vdash t = t, \Delta$ is valid. We take these as axioms.

Lemma: $\Gamma, t = u \vdash \phi[t], \Delta$ is valid iff $\Gamma, t = u \vdash \phi[u], \Delta$ is valid.

Here are the rules for the Hilbert symbol.

Lemma: For any term t, $\Gamma \vdash \Delta$ is valid iff $\Gamma, \phi[(\epsilon x.\phi)/x] \vdash \Delta$ is valid and $\Gamma \vdash \phi[t/x], \Delta$ is valid. If the existential quantifier is defined in terms of the Hilbert symbol, its rule can be derived from this rule (and the rule for the universal quantifier from the rule for the existential quantifier). Note that the Cut Rule is actually a special case of this rule.

Lemma: $\Gamma, \phi[(\epsilon x.\psi[x])/x], (\forall x.\psi[x] \leftrightarrow \chi[x]) \vdash \phi[(\epsilon x.\chi[x])/x], \Delta$.

Here is a lemma about provability which follows from common features of all our rules.

Lemma: If $\Gamma \vdash \Delta$ is provable using our sequent rules then $\Gamma, \Gamma' \vdash \Delta, \Delta'$ is also provable using our sequent rules for any finite sets Γ', Δ'.

These rules correspond precisely to our proof strategies for proof of quantified goals and use of quantified hypotheses. Our definition of proofs as formal objects can be extended to first order logic proofs by adding these sequent rules.

We now prove a constellation of results which show that first order logic is *complete* (any valid sequent can be proved using the rules we have given) but also cast some doubt on just how strong first-order logic is.

Observation: The sets of formal terms and formulas are countably infinite. It is obvious that they have countably infinite subsets, so they are not finite. A quick way to see that they are just countably infinite is to observe that all our objects (formulas, terms, sequents, and proofs) are built from natural numbers by pairing and the construction of finite sets, and that finite sets and pairs of natural numbers can be implemented as natural numbers, as we showed above. So the sets of terms and formulas could be understood as infinite sets of natural numbers. The formulation we use is advantageous because it is clearly adaptable to larger languages (we might for example want uncountably many constants). This argument also adapts to larger languages: for any set of an infinite cardinality κ, objects in the set can be used to code pairs of objects in the set by the theorem $\kappa^2 = \kappa$ of cardinal arithmetic, so if we for example have κ constants and otherwise the usual finite or countable complement of symbols we will have formula and term sets of size κ.

It is also important to note that the Construction which follows is valid for restricted languages. Limiting the number of constants, predicates, functions, relations and/or operators to a finite set does not affect the construction. Completely eliminating the Hilbert symbol does not affect the Construction. Using just one type or a finite subset of the types does not affect the Construction.

Construction: Let Γ and Δ be possibly infinite sets of formulas with the property that for any finite $\Gamma_0 \subseteq \Gamma$ and $\Delta_0 \subseteq \Delta$, $\Gamma_0 \vdash \Delta_0$ is not provable, and which are such that infinitely many constants of each type do not appear in any formula in either of these sets (this is not an essential limitation: constants a_i used can be replaced with a_{2i} in each type, freeing up infinitely many constants). Then there is a countably infinite structure in which each formula in Γ and the negation of each formula in Δ is satisfied.

For purposes of this proof we use only negation, disjunction, and the existential quantifier as logical operations (all the others are definable in terms of these and their proof rules are derivable from the rules for these and their definitions).

The fact that the model constructed will be countably infinite will be evident, because the elements of the model will be terms.

We provide an enumeration F_i of all the formulas of our language in which no bound variable appears free (every bound variable is in the scope of a quantifier over that variable), and in which shorter formulas appear before longer formulas.

We define sequences of finite sets of formulas Γ_i and Δ_i which will have the following properties.

1. Each $\Gamma_i, \Gamma' \vdash \Delta_i, \Delta'$ is not a provable sequent for any finite subsets Γ', Δ' of Γ, Δ respectively.
2. $\Gamma_i \subseteq \Gamma_{i+1}$; $\Delta_i \subseteq \Delta_{i+1}$
3. Each formula F_i appears in $\Gamma_{i+1} \cup \Delta_{i+1}$

The motivation is that the set Γ_∞ which is the union of all the Γ_i's will be the set of true statements of the model to be constructed and the set Δ_∞ which is the union of all the Δ_i's will be the set of false statements of the model to be constructed.

$\Gamma_0 = \Delta_0 = \emptyset$. The conditions are clearly satisfied so far.

If Γ_i and Δ_i are defined and the conditions are supposed satisfied so far, we have next to consider where to put F_i.

1. If $\Gamma_i, \Gamma' \vdash F_i, \Delta'$ is not provable for any finite subsets Γ', Δ' of Γ, Δ respectively, set $\Gamma_{i+1} = \Gamma_i$ and $\Delta_{i+1} = \Delta_i \cup \{F_i\}$.

2. If $\Gamma_i, \Gamma' \vdash F_i, \Delta'$ is provable for some $\Gamma' \subseteq \Gamma$ and $\Delta' \subseteq \Delta$, then it cannot be the case for any finite subsets Γ'', Δ'' of Γ, Δ respectively that $\Gamma_i, \Gamma'', F_i \vdash \Delta''$ is provable, as we would then be able to prove $\Gamma', \Gamma'' \vdash \Delta', \Delta''$ using the Cut Rule. If F_i is not of the form $(\exists x.\phi[x])$, we define Γ_{i+1} as $\Gamma_i \cup \{F_i\}$ and $\Delta_{i+1} = \Delta_i$. If F_i is of the form $(\exists x.\phi[x])$, let a be the first constant of the same type as x which does not appear in any formula in Γ, Δ, Γ_i or Δ_i, let Γ_{i+1} be defined as $\Gamma_i \cup \{(\exists x.\phi[x]), \phi[a]\}$ and let Δ_{i+1} be defined as Δ_i [an important alternative is to use the Hilbert symbol $(\epsilon x.\phi[x])$ instead of a]. Note that if $\Gamma_i, (\exists x.\phi[x]), \phi[a], \Gamma' \vdash \Delta_i, \Delta'$ were provable, so would $\Gamma_i, (\exists x.\phi[x]), \Gamma' \vdash \Delta_i, \Delta'$, and we have already pointed out that the latter cannot be proved for any subsets Γ', Δ' of Γ, Δ respectively in this case. [If the alternative approach is used, note that if $\Gamma_i, (\exists x.\phi[x]), \phi[(\epsilon x.\phi[x])/x], \Gamma' \vdash \Delta_i, \Delta'$ were provable, then $\Gamma_i, (\exists x.\phi[x]), \Gamma' \vdash \Delta_i, \Delta'$ would also be provable].

The discussion shows that the conditions required continue to hold at each stage of the construction. So the definition succeeds and we obtain sets Γ_∞ and Δ_∞ whose union is the set of all formulas and whose properties we now investigate.

We are able to show that the following Lemmas hold.

Lemma: Γ_∞ and Δ_∞ are disjoint.

Proof: If they had a common element A, then some Γ_i and Δ_i would have that common element, and $\Gamma_i \vdash \Delta_i$ would be an axiom of sequent calculus.

Lemma: $\Gamma \subseteq \Gamma_\infty$; $\Delta \subseteq \Delta_\infty$

Proof: Consider what happens to F_i in either of these sets at the appropriate stage of the Construction.

Lemma: $\neg\phi \in \Gamma_\infty \leftrightarrow \phi \in \Delta_\infty$; equivalently, $\neg\phi \in \Gamma_\infty$ iff ϕ is not in Γ_∞.

Proof: Otherwise for some i, Γ_i would contain both ϕ and $\neg\phi$ or Δ_i would contain both ϕ and $\neg\phi$. In either case $\Gamma_i \vdash \Delta_i$ would be provable.

Lemma: $\phi \vee \psi \in \Gamma_\infty$ iff either $\phi \in \Gamma_\infty$ or $\psi \in \Gamma_\infty$.

Proof: Otherwise we would either have $\phi \vee \psi$ in Γ_∞ and both ϕ and ψ in Δ_∞, in which case

$$\Gamma_i \vdash \Delta_i$$

for some i would take the form $\Gamma_i, \phi \vee \psi \vdash \phi, \psi, \Delta_i$, which would be provable, or we would have $\phi \vee \psi \in \Gamma_\infty$ and either $\phi \in \Delta_\infty$ or $\psi \in \Delta_\infty$, and thus some $\Gamma_i \vdash \Delta_i$ would take one of the forms

$$\Gamma_i, \phi \vdash \phi \vee \psi, \Delta_i$$

or

$$\Gamma_i, \psi \vdash \phi \vee \psi, \Delta_i,$$

both of which are provable.

Lemma: $(\exists x.\phi[x]) \in \Gamma_\infty$ iff there is a term t such that $\phi[t] \in \Gamma_\infty$.

Proof: If $(\exists x.\phi[x]) = F_i$ and $(\exists x.\phi[x]) \in \Gamma_\infty$ then some $\phi[a]$ is also in Γ_∞ by a specific provision of the construction. If $(\exists x.\phi[x]) \in \Delta_\infty$ and there is some $\phi[t] \in \Gamma_\infty$, then some $\Gamma_i \vdash \Delta_i$ takes the form

$$\Gamma_i, \phi[t] \vdash (\exists x.\phi[x]), \Delta_i$$

and this is provable.

Lemma: $t = t \in \Gamma_\infty$ for any term t. If $t = u \in \Gamma_\infty$ and $\phi[t] \in \Gamma_\infty$ then $\phi[u] \in \Gamma_\infty$.

Proof: Immediate from the form of the sequent rules for equality.

Lemma: The relation $=_n$ on terms of type n which holds between terms t and u of type n just in case $t = u \in \Gamma_\infty$ is an equivalence relation.

Proof: $t = u \vdash u = t$ and $t = u, u = v \vdash t = v$ are provable.

Lemma: For any term t, if $\phi[t/x] \in \Gamma_\infty$ then $\phi[(\epsilon x.\phi[x])/x] \in \Gamma_\infty$.

Proof: $\phi[t/x] \vdash \phi[(\epsilon x.\phi[x])/x]$ is provable.

Lemma: If $(\forall x.\phi[x] \leftrightarrow \psi(x)) \in \Gamma_\infty$ then $(\epsilon x.\phi[x]) = (\epsilon x.\psi[x]) \in \Gamma_\infty$.

Now we can define the interpretation of our language that we want. The elements of D_n are the terms of type n in our language. a_i^n is actually defined as $a_i^{\mathbf{n}}$ (each constant is its own referent). F_i^n is the map which sends each type n term t to the term $F_i(t)$. O_i^n sends each pair of type n terms $\langle t, u\rangle$ to the term $t\, O_i\, u$. P_i^n is the set of all terms t of type n such that $P_i(t) \in \Gamma_\infty$. R_i^n is the set of all pairs of type n terms $\langle t, u\rangle$ such that $t\, R_i\, u \in \Gamma_\infty$. The functions H_n are chosen so that $H_n(\{t \mid \phi[t/x] \in \Gamma_\infty\})$ is the formal term $(\epsilon x.\phi)$.

The idea here is that we construct a model in which each term is taken to represent itself. The atomic formulas are evaluated in a way consistent with the idea that $\phi \in \Gamma_\infty$ simply means "ϕ is true in the term model", and the lemmas above show that complex terms and formulas are evaluated exactly as they should be for this to work. We conclude that for each formula $\phi \in \Gamma$, ϕ is satisfied in the term model, and for each formula $\phi \in \Delta$, ϕ is not satisfied ($\neg\phi$ is satisfied) in the term model.

Definition: For any environment E whose range consists of closed terms and term or proposition T, we define $T[E]$ as $T[E(x_1)/x_1][E(x_2)/x_2]\ldots[E(x_n)/x_n]$ where n is the largest index of a variable which occurs free in T.

Theorem: In the interpretation of our language just described, $\mathcal{V}(E, \phi) = 1 \leftrightarrow \phi[E] \in \Gamma_\infty$ for each formal sentence ϕ, and $\mathcal{R}(E, t) = t[E]$ for each formal term t.

Indication of Proof: This is proved by induction on the structure of formal terms and propositions. The Lemmas above provide the key steps.

The following theorems follow from considering the Construction and following Theorem.

Completeness Theorem: Any valid sequent has a proof.

Proof: This is equivalent to the assertion that any sequent which is not provable is invalid. A sequent $\Gamma \vdash \Delta$ is invalid precisely if there is an interpretation of the language under which Γ consists entirely of true statements and Δ consists entirely of false statements. The Construction shows us how to do this for any sequent which cannot be proved.

Definition: A collection Γ of sentences is *consistent* iff there is an interpretation under which all of them are true.

Compactness Theorem: Any collection of sentences any finite subcollection of which is consistent is consistent.

Proof: Let Γ be a collection of sentences any finite subcollection of which is consistent. This implies that $\Gamma_0 \vdash \emptyset$ is invalid for each finite $\Gamma_0 \subseteq \Gamma$. This means that $\Gamma \vdash \emptyset$ satisfies the conditions of the Construction so there is an interpretation in a term model under which all the sentences in Γ are true.

Löwenheim-Skolem Theorem: Any consistent set of sentences has a finite or countable model. If it has models of every finite size it has an infinite model.

Proof: Any consistent set of sentences satisfies the conditions of the Construction, and so has a term model, which is countable (or finite). If the theory has models of every finite size, it is consistent with the theory resulting if we adjoin new constants a_i indexed by the natural numbers with axioms $a_i \neq a_j$ for each $i \neq j$, by Compactness. A model of this extended theory will of course be infinite.

The relation $=^n$ on D_n implementing on each type n will not be the equality relation on D_n, but it will be an equivalence relation. We can convert any model in which equality is represented by a nontrivial equivalence relation into one in which the equality relation is represented by the true equality relation on each type by replacing model elements of type n with their equivalence classes (or representatives of their equivalence classes) under $=^n$.

If the logic is extended to support our type theory, equality is definable. The relation $(\forall z.x \in z \rightarrow y \in z)$ provably has the properties of equality in the presence of the axiom of comprehension. Unfortunately, as we will see in the next section, full type theory does not satisfy the Completeness Theorem.

Although the set-theoretical definition of the Hilbert symbol involves Choice (and if we add type theory as part of our logic without some care, the properties of the Hilbert symbol will imply Choice) the Hilbert symbol adds no strength to first-order logic. If we have any theory not using the Hilbert symbol, we can use the Construction (without Hilbert symbols) to build an interpretation of the language of the theory in which all sentences

are evaluated, and then (since the domain of this interpretation is countable), add the order $t \leq u$ on terms defined by "the first term equal to t in the interpretation appears no later than the first term equal to u in the interpretation in a given fixed order on terms". Then define $(\epsilon x.\phi[x])$ as the first object in this order such that ϕ. The definition of $\leq$ extends to the new Hilbert terms, and all formulas involving the defined Hilbert symbol have valuations determined in the interpretation.

The alternative version of the Construction in which existential statements are witnessed by Hilbert symbols instead of new constants has the immediate merit that one does not need infinitely many free constants and the additional merit that every object in the term model is definable from the basic concepts of the theory (in the original version of the Construction, the witnesses have an anonymous quality).

If our language is made larger by providing an uncountable collection of constants, predicates, and/or function symbols, say of uncountable size κ, the Construction still works, with the modification that "$\Gamma \vdash \Delta$ is provable" should systematically be read "for some finite $\Gamma_0 \subseteq \Gamma$ and $\Delta_0 \subseteq \Delta$ $\Gamma_0 \vdash \Delta_0$ is provable". The difficulty is that the construction will pass through stages indexed by ordinals, and once $\alpha \geq \omega$ we will have Γ_α and Δ_α infinite sets. Note that we are not talking here about modifications which would make terms or formulas of the language themselves into infinite objects (such as infinite conjunctions or disjunctions). The Compactness Theorem is thus seen to hold for languages of all sizes, and likewise the Löwenheim-Skolem Theorem can be extended to assert that any theory with infinite models has models of each infinite size κ: to ensure that there are many distinct objects in a term model, add enough constants a_α with axioms $a_\alpha \neq a_\beta$ for each $\alpha \neq \beta$. Any finite collection of these new axioms will be consistent with any theory which has infinite models, and the Construction will give an interpretation under which all the new constants are distinct.

NOTE (everything to end of section):

Think about Omitting Types theorem here or later.

TNT is a nice exercise for this section. Also showing that type theory is distinct from Zermelo by showing that there are models of type theory with more natural numbers than types.

Section 6 is soon enough for development of the logic of the set constructor, but some allowance for the set constructor (and its type regime) should be added to syntax (which will require changes in my remarks). Add remarks about single-sorted theories being readily supported here, and more complex

multi-sorted theories possible but not needed.

5.4.1 Exercises

1. Express the axioms of group theory in the language of first order logic (you do not need types and you do not need to use numerical codings). Groups are exactly models of this theory. A group is said to have *torsion* if there is an element g of the group and a natural number n such that g^n is the identity element e of the group. A group is said to be *torsion-free* if it does not have torsion. Prove that there is no formula ϕ in our formal language for group theory which is true of exactly the groups with torsion. Hint: use compactness. Suppose that ϕ is a formula which is true in every group with torsion. Consider the sentences τ_n which say "there is a g such that $g^n = e$" for each concrete natural number n. Notice (explain) that each of these sentence can be written in our formal language. Verify that the infinite set of sentences $\{\phi, \neg\tau_1, \neg\tau_2, \neg\tau_3 \ldots\}$ satisfies the conditions of the Compactness Theorem (give details). Draw the appropriate conclusion.

   Explain why this tells us that $(\exists n \in \mathcal{N}.g^n = e)$ is not equivalent to any sentence in our formal language for group theory.

2. The Löwenheim-Skolem Theorem tells us that every theory with a finite or countable language has a finite or countable model. Our untyped set theory has a countably infinite language, so has countably infinite models.

   But in untyped set theory Cantor's Theorem $|A| < |\mathcal{P}(A)|$ holds. As an exercise in porting results from type theory to set theory, write out the proof of Cantor's Theorem in untyped set theory. Hint: you do not need to make finicky use of the singleton operator in your argument.

   Finally, if A is an infinite set in a model of untyped set theory, either A is not countably infinite (in which case we have an uncountable set) or A is countably infinite and $|A| < |\mathcal{P}(A)|$, in which case $\mathcal{P}(A)$ is an uncountable set (according to the model). Yet the whole model may be countably infinite, and so certainly any infinite subsets of the model are countably infinite. Why is this not a contradiction (this argument is called *Skolem's paradox*)? Hint: I'm using what look like the same words in different senses here; explain exactly how.

5.5 Cut Elimination for First-Order Logic

5.6 Incompleteness and Undefinability of Truth

We say that a term t is closed iff all bound variables appearing in it are actually bound by some quantifier (or Hilbert symbol). A closed formula in this sense is a sentence. Each closed term t has a referent which we may write $\mathcal{R}(t)$ (the choice of environment will not affect the reference of a closed term). There are terms 't' such that $\mathcal{R}(\text{'}t\text{'}) = t$: '$t$' has as its referent the formal term t itself. There is a recursive procedure (using our definition of syntax) which would allow us to define a function which sends every formal term t to such a formal term 't'. Similarly we can define a function sending each formal sentence p (considered as a mathematical object) to a formal term 'p' such that $\mathcal{R}(\text{'}p\text{'}) = p$.

An additional convention will make this easier to see: let the operator O_1 be reserved to represent the ordered pair, and the constants a_{2n} to represent the natural numbers n. Since all terms are built from natural numbers by pairing, easy recursive definitions of 't' in terms of t and 'p' in terms of p can be given.

Now we can prove some quite surprising theorems.

Gödel's First Incompleteness Theorem: There is a sentence of our language which is true but cannot be proved.

Proof: Define a predicate G of formulas p as follows: $G(p)$ says "p is a formula with one free variable x and $p[\text{'}p\text{'}/x]$ is not provable". We have seen in the previous sections that everything here is definable. Let g represent the formula $G(p)$ as a mathematical object. $G(g)$ says that g is a formula with one free variable (it has one free variable p as you can see above) and $g[\text{'}g\text{'}/p]$ is not provable. But $g[\text{'}g\text{'}/p]$ is the statement $G(g)$ itself. If $G(g)$ is true, it cannot be proved. If $G(g)$ is false, it can be proved and is therefore true. So $G(g)$ is true but not provable.

There are some subtleties here if there are unintended objects among our proofs (we discussed this possibility for the natural numbers earlier). The sentence $G(g)$ cannot be provable, as we would then have a concrete proof whose existence falsifies what it proves. Suppose that $G(g)$ could be decided by being proved false: this would show that there is a "proof" of $G(g)$, but that might be an "unintended object" that we would never actually find.

This loophole can be closed by modifying the definition of G (a trick due to Rosser). Instead of constructing a statement which asserts its own unprovability, construct by the same technique a statement which asserts that if it is provable there is a shorter proof of its negation (a notion of numerical measure of size of proofs can readily be defined recursively). If a concrete proof of this statement were given, there would be a proof of its negation which was shorter, and so also concrete. If a concrete disproof of this statement were given, then the statement would be true (as no shorter statement could be a proof): this would make a concrete proof of the statement possible. Whether or not there are unintended "proofs" or "disproofs" of this statement, the statement must actually be undecidable.

This theorem applies not only to our type theory but also to bounded Zermelo set theory, Zermelo set theory and *ZFC* (where all our constructions can be carried out) and even to arithmetic (our whole formal development of the notion of provability can be carried out entirely in arithmetic: all we need is a notion of ordered pair definable in arithmetic, and we have shown that enough set theory can be defined in arithmetic that Kuratowski pairs of natural numbers can be coded as natural numbers. Even our semantics can be defined in arithmetic, with the stipulation that environments have to be partial functions from variables to domain elements (since they must be finite) and domains D_n need to be defined by formulas rather than given as sets.

A corollary of Gödel's First Incompleteness Theorem is

Gödel's Second Incompleteness Theorem: Our type theory (or untyped set theory, or arithmetic) cannot prove its own consistency.

Indication of Proof: The underlying idea is that to prove consistency is to prove that some statements cannot be proved. If the Rosser sentence can be proved, we can prove that all sentences can be proved (because if the Rosser sentence has a proof, so does its negation, and so does everything). So if we can prove consistency we must be able to prove that the Rosser sentence cannot be proved. But if we can prove that the Rosser sentence cannot be proved, then we can prove that the Rosser sentence is (vacuously) true (and so we have proved it contrary to hypothesis).

There are problems of level here. To actually prove that all this works requires results such as "if we can prove ϕ, then we can prove that ϕ is provable," and some other similar proofs along the same lines.

We have never found the First Incompleteness Theorem particularly surprising: there was never any reason to suppose that we could prove everything that happens to be true in mathematics. The Second Incompleteness Theorem is a bit more alarming (we cannot prove that the reasoning techniques in our working theory are free from paradox *in that theory*). The next result is quite alarming (and requires more care to understand).

Tarski's Theorem: The predicate of formulas p of the language of our type theory (or of untyped set theory, or of arithmetic) which asserts that p is true cannot be defined in the same theory.

Proof: Suppose there there is such a definable predicate `true`. Define $T(p)$ as "p is a predicate with one free variable x and $\neg$`true`$(p[\text{`}p\text{'}/x])$". Let t be the mathematical object representing $T(p)$. Then $T(t)$ asserts that $T(t)$ itself is not true. This is simply impossible. There can be no truth predicate (of formal sentences).

It is easy to misunderstand this. For any statement ϕ in our informal mathematical language (of whichever theory) we can say "ϕ is true"; this simply means ϕ and has nothing to do with Tarski's theorem. What we cannot do is define a predicate of formal mathematical objects Φ coding sentences ϕ of the language of our working theory in such a way that this predicate is true of Φ exactly if the corresponding formula ϕ is true in our theory. This is quite weird, since the missing predicate can be understood as a predicate of natural numbers (in any of these theories, if we construe the pair of the formalization of syntax as the pair definable on the natural numbers).

The reader should notice the formal analogy between these results (especially Tarski's Theorem) and Russell's paradox. Unfortunately here the self-application $p[\text{`}p\text{'}/x]$ cannot be exorcised as $x \in x$ was by our type discipline: the self-application is meaningful so something else has to give.

It is important to notice that the problem here is not that our theories are too weak. Any theory sufficiently strong in expressive power to describe provability (which amounts to having enough arithmetic) has these features. It should be noted that stronger theories can prove consistency of weaker

theories. For example, type theory does prove the consistency of arithmetic (because one can build a set model of arithmetic in type theory).

6 Model Theory

NOTE: This should all be conducted in type theory.

6.1 Ultrafilters and Ultrapowers

Definition: Let $\leq$ be a partial order. A nonempty subset F of $\texttt{fld}(\leq)$ is a *filter in* $\leq$ iff it has the properties that for every $x, y \in \texttt{fld}(\leq)$ there is some z such that $z \leq x$ and $z \leq y$ and that for every x, y if $x \in F$ and $x \leq y$ implies $y \in F$. A filter in $\leq$ is proper iff it is not the entire field of F. A filter in $\geq$ is called an *ideal in* $\leq$.

Definition: This is a maximally abstract definition of filters and ideals. For our purposes in this section, the partial order $\leq$ will always be the subset relation on $\mathcal{P}(X)$ for some fixed set X. So, for the rest of this section, a filter on X is a subset of $\mathcal{P}(X)$ which is a filter in the subset relation on $\mathcal{P}(X)$ in the sense just defined. Further, an *ultrafilter* on X is a filter U on X with the property that for each $A \subseteq X$, exactly one of A and $X - A$ belongs to U. Note that for each $x \in X$, the set $U_x = \{A \in \mathcal{P}(X) \mid x \in A\}$ is an ultrafilter on X; such ultrafilters are called *principal* ultrafilters on X. An ultrafilter on X which is not of the form U_x for any $x \in X$ is called a *nonprincipal* ultrafilter on X.

Theorem: Let X be an infinite set. Then there is a nonprincipal ultrafilter on X.

Proof: Choose a well-ordering W of $\mathcal{P}(X)$. We define the ultrafilter U_W by transfinite recursion. Suppose that we have determined for each $\beta < \alpha$ whether $W_\beta \in U_W$. We provide that $W_\alpha \in U_W$ iff $W_\alpha \cap \bigcap_{\beta \in F} W_\beta$ is an infinite set for each finite set F of ordinals less than α such that $W_\beta \in U_W$ for each $\beta \in F$. Notice that the case $F = \emptyset$ tells us that W_α is infinite.

We verify that U_W is an ultrafilter on X.

The intersection of any finite subset of U_W is an infinite set: we can see this by considering the last element of the finite set in terms of the

well-ordering $\leq$ and applying the definition of U_W. A set A fails to belong to U_W exactly if there is a finite subcollection F of U_W such that the intersection of $F \cup \{A\}$ is finite: clearly if there is such a subcollection A is not in U_W, and if there is no such subcollection the recursive definition will place A in U_W.

We show that if A belongs to U_W and $A \subseteq B$, then B must belong to U_W: suppose B did not belong to U_W; it follows that there is a finite subcollection F of U_W such that the intersection of $F \cup \{B\}$ is finite, from which it follows that the intersection of $F \cup \{A\}$ is finite, from which it follows that A is not an element of U_W. We show that if A and B belong to U_W, there is $C \in U_W$ such that $C \subseteq A$ and $C \subseteq B$: a suitable C is $A \cap B$, for which it is clear that any finite subcollection F of U_W has the intersection of $F \cup \{A \cap B\}$ infinite because this is equal to the intersection of $(F \cup \{A\}) \cup \{B\}$. This verifies that U_W is a filter on X.

It cannot be the case that A and $X - A$ are both in U_W because their intersection is not infinite; nor can it be the case that both are not in U_W, because we would then have finite subsets F and G of U_W with the intersection of $F \cup \{A\}$ finite and the intersection of $G \cup \{X - A\}$ finite, so all but finitely many of the members of $\bigcap F$ would be outside A while all but finitely many of the members of $\bigcap G$ would be in A, so $\bigcap(F \cup G)$ would be finite, which is impossible. This verifies that U_W is an ultrafilter on X.

U_W is a nonprincipal ultrafilter because any principal ultrafilter U_x has a finite element $\{x\}$.

Note that the Axiom of Choice is used here (we have actually shown that there is a nonprincipal ultrafilter on X if $\mathcal{P}(X)$ can be well-ordered). This use of choice is essential: it is consistent with the other axioms of type theory or set theory that there is no nonprincipal ultrafilter on any infinite set. It is easy to show that any ultrafilter on a finite set is principal.

Definition: Let X be an infinite set and let U be a nonprincipal ultrafilter on X. Let A be any set (not necessarily of the same type as X). Let f and g be two maps from X to A (these may be lateral!). We define $f \sim_U g$ as holding iff $\{x \mid f(x) = g(x)\} \in U$. It is easy to see that $\sim_U$ is an equivalence relation: reflexivity and symmetry are

trivial, while transitivity follows from the fact that U is a filter: if $\{x \mid f(x) = g(x)\} \in U$ and $\{x \mid g(x) = h(x)\} \in U$, then $\{x \mid f(x) = g(x) \wedge g(x) = h(x)\} \in U$, being the intersection of two elements of U, and its superset $\{x \mid f(x) = h(x)\}$ is also in U. We define A^U, the *ultrapower* of A with respect to U, as the collection of equivalence classes under $\sim_U$. With each $a \in A$ we associate $a^* \in A^U$, defined as the equivalence class under $\sim_U$ of the constant function on X with value a. Note that the domain of $\sim_U$ is the collection of functions from X to A, and that we have indicated how to define this even if A and X are not of the same type.

Definition: Let X be an infinite set and let U be a nonprincipal ultrafilter on X. Let A and B be sets (not necessarily of the same type) and let R be a (possibly lateral) relation from A to B. For $[f]$ in A^U and $[g]$ in B^U, we define $[f]\, R^U\, [g]$ as holding iff $\{x \mid f(x)\, R\, g(x)\} \in U$ (it is straightforward to show that this does not depend on the choice of the representatives f and g of the elements of A^U and B^U). Note that $a^*\, R^U\, b^* \leftrightarrow a\, R\, b$.

Construction: We view A^U as a kind of extension of A, with each element a of A corresponding to the element a^* of A^U. We are going to define an extension of the language we use to talk about A to a language which talks about A^U. In fact, we are going to carry out such an extension for any collection of domains we wish to consider, all at once.

For any open sentence $\phi(x_1, \ldots, x_n)$ with no free variables other than $x_1, \ldots, x_n$, in which each $x_i \in A_i$, we define a sentence $\phi^*([f_1], \ldots, [f_n])$ for any fixed $[f_i] \in A_i^U$ as meaning $\{x \mid \phi(f_1(x), \ldots, f_n(x))\} \in U$.

If $f_i \equiv_U g_i$ for each i, then $\phi^*([f_1], \ldots, [f_n])$ asserts that $\{x \mid \phi(f_1(x), \ldots, f_n(x))\}$ is an element of U, and, because intersections of elements of U are in U, so is $\{x \mid \phi(f_1(x), \ldots, f_n(x)) \wedge f_1(x) = g_1(x) \wedge \ldots \wedge f_n(x) = g_n(x)\}$, which is a subset of $\{x \mid \phi(g_1(x), \ldots . g_n(x))\}$, so this latter set is in U, so $\phi^*([g_1], \ldots, [g_n])$. The argument is completely symmetrical that shows that $\phi^*([g_1], \ldots, [g_n])$ implies $\phi^*([f_1], \ldots, [f_n])$, so the choice of representatives in our notation for elements of A_i^U's is immaterial.

We note that if $\phi(x_1, \ldots, x_n)$ is $\neg\psi(x_1, \ldots, x_n)$, then $\phi^*([f_1], \ldots, [f_n])$ is equivalent to $\{x \mid \neg\psi(f_1(x), \ldots, f_n(x))\} \in U$, which is equivalent to $\{x \mid \psi(f_1(x), \ldots, f_n(x))\} \notin U$, because U is an ultrafilter, which is in

turn equivalent to $\neg\psi^*([f_1],\ldots,[f_n])$. In other words, the meaning of negation in the translated language is what we expect.

If $\phi(x_1,\ldots,x_n)$ is $\psi(x_{s_1},\ldots,x_{s_p})\wedge\chi(x_{t_1},\ldots,x_{t_q})$, then $\phi([f_1],\ldots,[f_n])$ is equivalent to $\{x \mid \psi(f_{s_1}(x),\ldots,f_{s_p}(x))\wedge\chi(f_{t_1}(x),\ldots,f_{t_q}(x))\}\in U$, which is equivalent to $\{x \mid \psi(f_{s_1}(x),\ldots,f_{s_p}(x))\}\in U\wedge\{x \mid \chi(f_{t_1}(x),\ldots,f_{t_q}(x))\}\in U$, because subsets A and B of X both belong to U iff their intersection belongs to U, and this is in turn equivalent to $\psi^*([f_{s_1}],\ldots,[f_{s_p}])\wedge\chi^*([f_{t_1}],\ldots,[f_{t_q}])$. In other words, the meaning of conjunction in the translated language is what we expect.

If $\phi(x_1,\ldots,x_n)$ is $(\exists y.\psi(y,x_1,\ldots,x_n))$, then $\phi^*([f_1],\ldots,[f_n])$ is equivalent to $\{x \mid (\exists y.\psi(y,f_1(x),\ldots,f_n(x))\}\in U$. If there is a g such that $\{x \mid \psi(g(x),f_1(x),\ldots,f_n(x))\}\in U$, then certainly $\{x \mid (\exists y.\psi(y,f_1(x),\ldots,f_n(x))\}\in U$, because $\{x \mid \psi(g(x),f_1(x),\ldots,f_n(x))\}\subseteq\{x \mid (\exists y.\psi(y,f_1(x),\ldots,f_n(x))\}$. Now suppose that $\{x \mid (\exists y.\psi(y,f_1(x),\ldots,f_n(x))\}\in U$. Define a function g such that for each x such that $(\exists y.\psi(y,f_1(x),\ldots,f_n(x))$ we have $\psi(g(x),f_1(x),\ldots,f_n(x))$: this is an application of the Axiom of Choice. Now we have $\{x \mid \psi(g(x),f_1(x),\ldots,f_n(x))\}=\{x \mid (\exists y.\psi(y,f_1(x),\ldots,f_n(x))\}\in U$ for this particular g. So we have shown that $\phi^*([f_1],\ldots,[f_n])$ iff there is a $[g]$ such that $\psi^*([g],[f_1],\ldots,[f_n])$. This means that the existential quantifier over any A_i in the base language translates to the existential quantifier over A_i^U in the extended language (here we moved the quantified argument into first position, but it should be clear that we do not really lose any generality by doing this).

Note that if $\phi(x_1,\ldots,x_n)$ is $\psi(a,x_1,\ldots,x_n)$, then $\phi^*([f_1],\ldots,[f_n])$ is equivalent to $\{x \mid \psi(a,f_1(x),\ldots,f_n(x))\}\in U$, which is equivalent to $\psi^*(a^*,[f_1],\ldots,[f_n])$, which indicates that constants taken from domains A_i behave naturally in the extended language.

In the last two paragraphs, we have done manipulations on the first argument of an open sentence which can of course be done on any argument; since we can change the indexing of the arguments (and so of the domains) of a fixed open sentence it should be clear that we do not lose generality.

Note finally that if $\phi(x_1,\ldots,x_n)$ is true for any assignment of values to the x_i's from the appropriate A_i's, then $\{x \mid \phi(f_1(x),\ldots,f_n(x))\}=X\in U$ for any choice of f_i's, so $\phi^*([f_1],\ldots,[f_n])$ is always true. Transla-

tions of general truths about the A_i's hold true in the extended language over the A_i^U's.

6.2 Technical Methods for Consistency and Independence Proofs

There is a political point to be made here: all of these things can be done in type theory, quite naturally, and can thence be exported to *NFU* without reference to the usual set theory.

6.2.1 Frankel-Mostowski Methods; The Independence of Choice

6.2.2 Constructibility and the Minimal Model of Type Theory

Build the Forster term model of type theory. Also, prove the consistency of CH and GCH (though this might get forced forward after the logic section, because there is model theory involved.).

6.2.3 Forcing and the Independence of CH

The treatment of constructibility in the previous subsection is precisely that in the usual set theory (the fact that all the work is done in Z should make this clear. Our treatment of forcing is somewhat different from the treatement in the usual set theory: this can be seen from the fact that it handles atoms, which the usual techniques do not, and also from the fact that it *creates* atoms. The differences are technical: the basic idea is the same. What we do show by this method is that it appears that it is not necessary to do recursion along the cumulative hierarchy to do forcing (as is commonly done).

6.2.4 Generalizing the T operation

NOTE: this note might better belong somewhere else, but these considerations are needed here.

Certain collections, such as the natural numbers, are "the same" in each sufficiently high type. This is usually witnessed by a T operation. Some collections on which a T operation is defined get larger at each type; these are of less interest to us here.

T operations are defined on cardinals and on ordinals (more generally on isomorphism types) already. We point out that if we have defined T operations on sets A and B, there is a natural way to define a T operation on $\mathcal{P}(A)$ (for $a \subseteq A$, define $T^{\mathcal{P}(A)}(a)$ as $T^A\text{“}a$), on B^A (so that $T^{B^A}(f)(T^A(a)) = T^B(f(a))$, and on $A \times B$ (so that $T^{A\times B}(\langle a, b\rangle) = \left\langle T^A(a), T^B(b)\right\rangle$). We superscript T operations with their intended domains here for precision: we will not usually do this.

There is a uniform way to define T operations on sets with a certain kind of symmetry.

Definition: We call a bijection $f : V \to V$ a *permutation of the universe*. We use Π as a nonce notation for the set of all permutations of the universe. Define $j(f)$ so that $j(f)(x) = f\text{“}x$ for all x ($j(f)$ is undefined on sets with urelements as members). Define $j^{\mathbf{n}}(f)$ in the obvious way. Further, we define the operation $j^{\mathbf{n}}(\iota)$ similarly (with due respect to the fact that ι is itself a type-raising operation, but the definition works formally). A set A is *n-symmetric* iff $j^{\mathbf{n}}(f)(A) = A$ for all permutations of the universe f of the appropriate type. Notice that this implies that $A \in \mathcal{P}^{\mathbf{n}}(V)$. We define a T operation on n-symmetric objects A for each n:

$$T(A) = \{j^{\mathbf{n-1}}(f)(j^{\mathbf{n-1}}(\iota)(a)) \mid a \in A \wedge f \in \Pi\}.$$

Observation: The generalized T operation here would coincide with all T operations defined up to this point, if we used the Kuratowski ordered pair, or if we presumed that the type-level ordered pair coincided with the Quine ordered pair on sets and restricted all use of pairing to sets of sets (as would happen if we assumed strong extensionality). For cardinal numbers are 2-symmetric, isomorphism types are 4-symmetric if defined in terms of Kuratowski pairs and 2-symmetric if defined in terms of Quine pairs, and the definitions given above for power sets, function spaces, and cartesian products will coincide with appropriate T operations of this kind on power sets, function spaces and cartesian products (taking into account the effect on the degree of symmetry of these set constructions).

6.2.5 Forcing: Basic Definitions

We fix a definable partial order $\leq_P$ with field P which supports a T operation with the property that $T\text{“}(\leq_P) = \leq_P$ (which of course implies that $T\text{“}P = P$). This is of course a pun: what is being said is that the definition of P with all types raised by one will give the image under the T operation of the original partial order P. Such an order P will be defined and essentially "the same" structure in all types above a certain level.

The set P will be in some sense the space of "truth values" for the forcing interpretation. Each element of $\leq_P$ represents an (incomplete) "state of information"; the relation $p \leq_P q$ tells us that the state of information described by q extends the state of information described by p (the opposite convention is often used!). If neither $p \leq_P q$ nor $q \leq_P p$, the states of information described by p and q are to be understood to be incompatible.

"The objects of type n" of our forcing interpretation are relations x from $V^{\mathbf{n}}$ to P, that is, subsets of $V^{\mathbf{n}} \times P$, with the property that $\langle y, p\rangle \in x \wedge q \geq_P p \to \langle y, q\rangle$. Notice that the type n objects of the forcing model are actually certain type $n+1$ objects. The type $n+1$ objects which will be interpreted as type n objects are called *names*. Those familiar with treatments of forcing in the usual set theory should notice that we are *not* requiring names to be relations from *names* to elements of P: this would introduce a recursion on the type structure, which is something always to be avoided in type theory. We will see below how difficulties which might be supposed to arise from this freedom in the construction of names are avoided.

The central definition of the forcing interpretation is the definition of a notation $p \vdash \phi$ for formulas ϕ of type theory, which is intended to tell us when a condition p gives us sufficient information to decide that an assertion ϕ is true.

The central theorem of the forcing interpretation will be that $p \vdash \phi$ is true for each axiom ϕ, that $p \vdash \phi$ can be deduced from $p \vdash \psi$ whenever ϕ can be deduced from ψ by a rule of logic. It will further be clear that we cannot prove $\neg(p \vdash \phi \wedge \neg\phi)$ (unless we can prove a contradiction in type theory itself). It is very important to notice that this is not metamathematics: $p \vdash \phi$ is not an assertion about a

mathematical object 'ϕ' coding the assertion ϕ as in the development of Gödel's theorem or Tarski's theorem, and we are not building a set model of type theory (this cannot be done in type theory by those very theorems!). Of course we may associate with set models of type theory (if there are any) set models of type theory generated by applying a forcing interpretation to those set models, and this will be of some interest.

Definition: We define

$$\mathbb{N}_P = \{x \in \mathcal{P}(V \times P) \mid (\forall y.(\forall p \in P.(\forall q \geq_P p.\, \langle y, p\rangle \in x \rightarrow \langle y, q\rangle \in x)))\}$$

as the set of *P-names*. We define the notation $p \vdash \phi$ recursively. We suppose all logical operators defined in terms of $\wedge, \neg, \forall$.

negation: $p \vdash \neg\phi$ is defined as $(\forall q \geq_P p.\neg(q \vdash \phi))$. Informally, "no matter how much information we add to p, we will not verify ϕ".

conjunction: $p \vdash \phi \wedge \psi$ is defined as $(p \vdash \phi) \wedge (p \vdash \psi)$. This appears simple enough, but one should note that if one expands out the definition of disjunction or implication in terms of the given definitions of negation and conjunction one does not get this nice distributivity.

universal quantification: $p \vdash (\forall x.\phi)$ is defined as $(\forall \mathbf{x} \in \mathbb{N}_P.p \vdash \phi[\mathbf{x}/x])$. Again, this definition looks very direct, but it is instructive to analyze the expansion of $p \vdash (\exists x.\phi[x])$.

pseudo-membership: (this will not be the interpretation of membership, for reasons that will become evident, but it makes the definition easier): for any x, y, $p \vdash x \in^* y$ iff $y \in \mathbb{N}_P \wedge (\forall q \geq_P T^{-1}(p).(\exists r \geq_P q.\, \langle x, r\rangle \in y))$. Note the necessity of the introduction of the T operator so that we have a well-formed assertion of type theory. Note also that x here is any object at all (of appropriate type) while y is a name of the next higher type.

Pseudo-membership does not officially appear in formulas of our language; this notation is used only in the definitions of equality and membership for the forcing interpretation.

equality: Let x and y be names. $p \vdash x = y$ is defined as

$$(\forall z.(p \vdash z \in^* x) \leftrightarrow (p \vdash z \in^* y)).$$

Names are asserted to be equal as soon as we have enough information to see that they have the same pseudo-members.

sethood: $p \vdash \mathtt{set}(x)$ is defined as

$$(\forall y.(p \vdash y \in^* x) \to y \in \mathbb{N}_P \wedge (\forall z.(p \vdash y = z) \to (p \vdash z \in^* x))).$$

p says that x is a set iff anything that p thinks is a pseudo-element of x is a name and any name that p thinks is equal to an pseudo-element of x p also thinks is an pseudo-element of x. We will see that under these conditions we can drop the "pseudo-".

membership: $p \vdash x \in y$ is defined as $(p \vdash x \in^* y) \wedge (p \vdash \mathtt{set}(y))$. The idea here is that we convert the names whose pseudo-extension does not respect equality to urelements. This is how we avoid recursion on type in our definitions (along with the fact that we use typically ambiguous partial orders on forcing conditions).

type-shifting convention: Notice that in atomic formulas we have p at the same type as the highest type of one of the arguments. Hereafter we stipulate $p \vdash \phi$ iff $T(p) \vdash \phi$; the type of p may freely be shifted. It would otherwise be difficult to type conjunctions, and it should be clear that this will introduce no conflicts.

NOTE: in the context of NF(U) this will be clear if the set P is strongly cantorian. What can be done (if anything) with cantorian partial orders needs to be cleared up [when it is cleared up the exact way we proceed here might need to be modified].

7 Saving the Universe: Stratified Set Theories

This section concerns a class of untyped set theories which are related to type theory (as Zermelo set theory and *ZFC* also are) but in a different way. The first theory of this class was introduced by Quine in his "New foundations for mathematical logic" (1937) and so is called *NF*, which is short for "New Foundations". *NF*, as we shall see, is a very strange theory

for rather unexpected reasons. We shall ignore historical precedent and start by introducing *NFU* (New Foundations with urelements), which is much more tractable. *NFU* was shown to be consistent by R. B. Jensen in 1969.

Most of the theories of this class share the perhaps alarming characteristic that they assert the existence of a universal set.

7.1 Introducing *NFU*

The starting point of the line of thought which led Quine to "New Foundations" but which will lead us first to *NFU* (due to careful planning) is an observation which we have already exploited. The types of our type theory are very similar to one another (in terms of what we can prove). We have used this observation to avoid cluttering our notation with endless type indices. We begin by carefully stating the facts already known to us (at least implicitly) about this ambiguity of type and considering some extrapolations.

7.1.1 Typical Ambiguity Examined

If we suppose that each variable x in the language of our type theory actually comes with a type index ($x^{\mathbf{n}}$ is the shape of the typical type n variable), we can define an operation on variables: if x is a variable of type n, we define x^+ as the variable of type $n+1$ obtained by incrementing the type index which x is supposed to have (though we continue our convention of not expressing it). This allows us to define an operation on formulas: if ϕ is a formula of the language of type theory, we define ϕ^+ as the result of replacing every variable x (free or bound) in ϕ with the type-incremented x^+. The same operation can be applied to terms: $\{x \mid \phi\}^+ = \{x^+ \mid \phi^+\}$, and $(\epsilon x.\phi)^+ = (\epsilon x^+.\phi^+)$.

Our first observation is that for any formula ϕ, ϕ^+ is also a formula, and for any term T, T^+ is also a formula. The converse is also true. Further, if ϕ is an axiom, ϕ^+ is also an axiom (in fact, the converse is also true). Further, if ψ can be deduced from ϕ by any logical rule, ψ^+ can also be deduced from ϕ^+, whence it follows that if ϕ is a theorem of type theory, ϕ^+ is also a theorem of type theory. In this case, the converse is not necessarily the case, though the converse does hold in *TNT*. This means that anything we know about a particular type (and a number of its successors) is also true in each higher type (and a number of its corresponding, appropriately type-shifted successors). Further, any object we can construct in type theory has a correlate constructed in the same way at each higher type. We have exploited this

phenomenon, which Whitehead and Russell called "systematic ambiguity" in the more complex system of their *Principia Mathematica*, which most workers in the area of *NF* now call "typical ambiguity", and which is a rather extreme example of what computer scientists call *polymorphism*, to make it almost completely unnecessary to mention specific type indices in the first section of this book.

Quine made a daring proposal in the context of a type theory similar to ours (in fact, differing only in the assumption of strong extensionality). He suggested that it is not just the case that provable statements are the same at each type, but that the same statements are true in each type, and that the objects at the different types with correlated definitions do not merely serve as the subjects of parallel theorems but are in fact the same objects. The theory which results if this proposal is applied to our type theory is an untyped set theory, but rather different from the theory of Zermelo developed above.

In this theory we have not a universal set $V^{\mathbf{n+1}} = \{x^{\mathbf{n}} \mid x^{\mathbf{n}} = x^{\mathbf{n}}\}$ for each n, but a single set $V = \{x \mid x = x\}$. We have already shown that it follows from the Axiom of Separation of Zermelo set theory that there can be no such set V (whence it follows that if this new theory is coherent it does not satisfy the Axiom of Separation). We do not have a $3^{\mathbf{n+1}}$ which contains all the three-element sets of type n objects, but a single object 3 which is the set of all three-element sets.

We will give the precise definition of this theory in the next section. What we will do now is prove a theorem due to Specker which will make the connections between various forms of typical ambiguity clearer. For the rest of this section, we discuss theories framed in languages in which variables are typed and which satisfy the condition that for any formula ϕ is well-formed if and only if ϕ^+ is well-formed. Further, we require that the language of the theory be closed under the basic logical operations familiar to us from above, and that whenever the rules allow us to deduce ϕ from ψ [neither formula mentioning any maximum type] we are also able to deduce ϕ^+ from ψ^+. It is required that every context in which a term can occur dictates the type of that term exactly.

We consider the following suite of axioms.

Ambiguity Scheme: For each sentence ϕ (formula with no free variables) for which ϕ^+ is well-formed, $\phi \leftrightarrow \phi^+$

With any theory T in typed language, we associate a theory T^∞ whose

sentences are simply the sentences of T with all type distinctions removed. A model of T^∞, if there is one, is a model of the typed theory T in which all the types are actually the same. Notice that T^∞ is automatically the same as $(T + Amb)^\infty$, where Amb is the ambiguity scheme above, because Amb^∞ is a set of tautologies.

Note that the language of T^∞ allows things to be said which cannot be said in the typed language of T: sentences like $a \in a$ are well-formed, and a completion of a consistent T^∞ would assign truth values to such sentences.

Theorem (Specker): For any theory in typed language which is well-behaved in the ways outlined above, T^∞ is consistent iff $T + Amb$ is consistent.

Proof: It is obvious that the consistency of T^∞ implies the consistency of $T + Amb$.

Suppose that $T + Amb$ is consistent. Our goal is to show that T^∞ has a model. We first observe that this is obvious if the language of T contains the Hilbert symbol (or any construction with equivalent logical properties). For $T + Amb$, being consistent, can be extended to a complete theory, which has a model consisting entirely of closed terms T built using the Hilbert symbol. We can then identify the term T with the term T^+ for every T. No conflict can occur: any assertion $\phi(T)$ has the same truth value as $\phi^+(T^+)$ (and these identifications and equivalences can be indefinitely iterated) [and no weird variants such as $\phi^+(T)$ are meaningful]. The truth value of $\phi(a_1, \ldots, a_n)$ with any Hilbert symbol arguments a_i however weirdly typed can be established by raising the type of ϕ sufficiently high that the types expected for its arguments are higher than the types of any of the a_i's then raising the types of the arguments a_i to the correct types, then evaluating this well-typed formula.

To complete the proof we need to show that any typed theory $T + Amb$ can be extended to include a Hilbert symbol in a way which preserves the truth of all sentences and allows Amb to be extended to the new sentences. Since $T + Amb$ is consistent, we can suppose it complete. We list all Hilbert symbols, stipulating that a Hilbert symbol must appear after any Hilbert symbol which occurs as a subterm of it in the list. We assume that before each Hilbert symbol is introduced we have a deductively closed theory which contains all instances of Amb appropriate to its language (i.e., not instances which mention Hilbert symbols not yet

introduced). We introduce the Hilbert symbol $a = (\epsilon x.\chi[x])$. We then find a maximal collection of sentences $\phi[a]$ which includes $\chi[a]$, contains all type-raised copies of its elements, and is consistent. We can do this by finding a maximal collection of sentences $\phi[a]$ which is consistent: for $(\exists x.\Phi[x])$ will be consistent for all conjunctions $\Phi[x]$ of elements of the set, and by *Amb* $(\exists x.\Phi[x])^{+^i}$ will be consistent for each i: thus we can assume not only each $\phi[a]$ in the set but each $\phi[a]^{+^i}$. We add all these sentences to our theory. We now assume that we have a complete set of sentences $\Phi[a, a^+, \ldots, a^{+^k}]$ consistent with our theory and closed under $+$ (we have just dealt with the base case $k = 1$). We show that we can get a complete set of sentences $\phi[a, a^+, \ldots, a^{+^{k+1}}]$ consistent with our theory and closed under $+$. Suppose $\psi[a, a^+, \ldots, a^{+^{k+1}}]$ is a sentence which we wish to consider. We consider the status of sentences $(*) : (\exists x.\psi[a, a^+, \ldots, a^{+^k}, x] \wedge \Phi^+[a^+, \ldots, a^{+^k}, x])$ and $(*^\neg) : (\exists x.\neg\psi[a, a^+, \ldots, a^{+^k}, x] \wedge \Phi^+[a^+, \ldots, a^{+^k}, x])$ which are already decided in our theory (because they mention blocks of k successive type-shifted versions of a). We see that if $\psi[a, a^+, \ldots, a^{+^k}, a^{+^{k+1}}]$ (resp. $\neg\psi[a, a^+, \ldots, a^{+^k}, a^{+^{k+1}}]$ is consistent with our theory then this statement must have already been decided as true (otherwise we would be able to disprove $\psi[a, a^+, \ldots, a^{+^k}, a^{+^{k+1}}]$ (resp. $\neg\psi[a, a^+, \ldots, a^{+^k}, a^{+^{k+1}}]$) from prior assumptions). This means that we can extend the sequence of k type shifted versions of a with a new term in such a way that the "type shifted sequence" starting with a^+ and extended with x has as many of the known properties of blocks of k type shifted versions of a as we want, and the sequence of $k + 1$ elements satisfies ψ (resp. $\neg\psi$). These properties can include the ability (expressed in the formula $(*)$ (resp. $(*^\neg)$ above, which can be used to extend Φ) to further extend the sequence as many times as desired, while also preserving the property that blocks of $k + 1$ elements of the extended sequence satisfy (type shifted versions of) ψ (resp. $\neg\psi$). Compactness then tells us that we can assume that all blocks of $k + 1$ type shifted versions of a satisfy ψ (resp. $\neg\psi$). This means that we can proceed (again by compactness) to find a maximal collection of consistent sentences $\psi[a, a^+, \ldots, a^{+^k}, a^{+^{k+1}}]$ such that the closure of this set under $+$ is consistent with our previous theory. Repeating this process for all k gives us a theory with the new Hilbert symbol adjoined which extends *Amb* as desired. Repeating this process for all Hilbert symbols gives the desired extension of $T + Amb$ with Hilbert symbols, and with the scheme

Amb extended appropriately to Hilbert symbols.

7.1.2 Definition and Consistency of *NFU*

We refer to the typed theory of sets which is our working theory as *TSTU* (excluding for the moment the axioms of Infinity, Ordered Pairs, and Choice). We refer to *TSTU* + strong extensionality as *TST*. We define *NFU* (for the moment) as $TSTU^{\infty}$, and define *NF* ("New Foundations") as TST^{∞}.

In this section we will expand a bit on how to understand the theory *NFU*, prove its consistency, and observe that the method of proof extends to a stronger theory which we will then make the referent of the name *NFU*.

NFU is an untyped set theory, like the theories of chapter 4. The axioms of *NFU* are exactly the axioms obtained from axioms of Extensionality and Comprehension of *TSTU* by disregarding all distinctions of type between the variables. Impossible axioms like $\{x \mid x \notin x\}$ do not appear as instances of Comprehension because $x \notin x$ is not the shape of any formula of the language of *TSTU*: we drop the type distinctions, but this does not introduce identifications between variables.

We recapitulate the axioms of *NFU*.

Primitive notion: There is a designated object $\emptyset$ called the *empty set*.

Axiom of the empty set: $(\forall x.x \notin \emptyset)$.

Definition: We say that an object x is a *set* iff $x = \emptyset \vee (\exists y.y \in x)$. We write $\mathtt{set}(x)$ to abbreviate "x is a set" in formulas. We say that objects which are not sets are *atoms* or *urelements*.

Axiom of extensionality:

$$(\forall xy.\mathtt{set}(x) \wedge \mathtt{set}(y) \rightarrow x = y \leftrightarrow (\forall z.z \in x \leftrightarrow z \in y)),$$

In these axioms, the only changes we make are complete omission of references to types and type indices. The comprehension axiom is trickier.

***Axiom of comprehension:** For any formula $A[x]$ obtained by ignoring type distinctions in a formula of the language of type theory in which the variable y (of type one higher than x) does not appear,

$$(\exists y.(\forall x.x \in y \leftrightarrow A[x])).$$

We star this because it is not the form of the axiom we will use.

Definition: A formula ϕ of the language of set theory is said to be "stratified" iff there is a function σ (called a *stratification* of ϕ) from variables to natural numbers (or, equivalently, integers) such that for each atomic formula $x = y$ appearing in ϕ we have $\sigma(x) = \sigma(y)$ and for each atomic formula $x \in y$ appearing in ϕ we have $\sigma(x) + 1 = \sigma(y)$. Note that for a formula in equality and membership alone, to be stratified is precisely equivalent to being obtainable from a formula of the language of type theory by ignoring type distinctions

Axiom of stratified comprehension: For any stratified formula $A[x]$ in which the variable y does not appear,

$$(\exists y.(\forall x.x \in y \leftrightarrow A[x])).$$

The axiom of extensionality tells us that there is only one such object y which is a set (there may be many such objects y if $A[x]$ is not true for any x, but only one of them ($\emptyset$) will be a set). This suggests a definition:

Set builder notation: For any stratified formula $A[x]$, define $\{x \mid A[x]\}$ as the unique *set* of all x such that $A[x]$: this exists by Comprehension and is uniquely determined by Extensionality.

We show that *NFU* is consistent. We have shown above that it suffices to demonstrate that *TSTU*+Amb is consistent.

Let Σ be any finite collection of sentences of the language of *TSTU*. Let n be chosen so that Σ mentions only types $0 - (n-1)$. Choose a sequence of sets X_i such that $|\mathcal{P}(X_i)| \leq |\iota\text{“}X_{i+1}|$ for each i. Choose injective maps $f_i : \mathcal{P}(X_i) \to \iota\text{“}X_{i+1}$ for each i and define relations $x \in_i y$ as $x \in X_i \wedge y \in X_{i+1} \wedge x \in f_i^{-1}(\{y\})$ (where of course this is understood to be false if $f_i^{-1}(\{y\})$ is undefined). It is easy to see that the resulting structure is a model of TSTU: the interpretation of a sentence of TSTU is obtained by replacing each type i variable with a variable restricted to X_i, and replacing each occurrence of $\in$ in an atomic formula $x \in y$ with $x \in_i y$, where i is the type of x. It should be easy to see that the interpretation of each axiom is true. Notice that this construction is carried out in our type theory, with the types of all the elements of the X_i's being the same fixed type whose identity does not matter for our purposes.

Now observe further that for any strictly increasing sequence s of natural numbers, the sequence X^s defined by $X^s_i = X_{s_i}$ determines an interpretation of TSTU in exactly the same way. We observe that the sentences Σ determine a partition of the n-element sets A of natural numbers as follows: consider a sequence s such that $s\text{“}\{0,\ldots,n-1\} = A$ and note the truth values of the sentences of Σ in the models X^s (which will be entirely determined by the first n terms of X^s). This is a partition of the n element subsets of $\mathbb{N}$ into no more than $2^{|\Sigma|}$ parts, which by Ramsey's theorem has an infinite homogeneous set H. Now consider any X^s such that $s\text{“}\mathbb{N} \subseteq H$: the interpretations of all sentences $\phi \leftrightarrow \phi^+$ for ϕ in the axiom scheme Amb will be true in such models. We have shown that every finite subset of Amb is consistent with TSTU, so by Compactness TSTU + Amb is consistent, so by Specker's theorem on ambiguity, *NFU* is consistent.

We have used more mathematical power than we need here. We have assumed in effect that $\beth_\omega$ exists (because we assume the existence of an infinite sequence X_i). This is not strictly necessary: we can use a more refined form of Ramsey's theorem and show the existence of homogeneous sets of sufficient size in sufficiently long finite sequences of X_i's. However, we do not regard the existence of $\beth_\omega$ as a dubious assumption.

The method of proof used here extends to any extension of *TSTU* with ambiguous axioms. For example *NFU* + Infinity + Choice is shown to be consistent by this argument. Further, we can add the axiom of Ordered Pairs as well: add predicates π_1 and π_2 with the additional rules that typing for formulas $x\,\pi_i\,y$ follows the same rules as typing for formulas $x = y$ and additional axioms $(\forall x.(\exists! y.x\pi_i y))$ (each π_i is a function of universal domain) and $(\forall xy.(\exists! z.z\pi_1 x \wedge z\pi_2 y))$. These axioms hold in our working theory, and can be made to hold in the X_i's by stipulating that each X_i is infinite and providing bijections $\Pi : (X_i \times X_i) \to X_i$ for each i, and interpreting $x\pi_j y$ between type i objects as holding iff $y = \pi_j(\Pi_i(x))$.

Hereinafter we will usually mean *NFU* + Ordered Pairs + Choice when we refer to NFU.

We further note that stratification can be extended to a language with terms, if a stratification must take the same value at $(\epsilon x.\phi)$ that it does at x (the structure of ϕ then dictating type differentials between x and any parameters in the term), and noting that any term construction can be supposed implemented by a Hilbert epsilon term. This can be handled in the consistency proof by fixing choice functions to identify referents of Hilbert epsilon terms in the X_i's.

This proof allows us to bootstrap our working theory from *TSTU* with Ordered Pairs and Choice to *NFU* with Ordered Pairs and Choice, if we are so inclined: we can adopt the view that the types of our theory, which are suspiciously similar because we have been careful to keep our methods of proof over them entirely uniform, are in fact all the same domain. We will explore the consequences of taking this perhaps odd view.

(NOTE: we certainly want to consider the Boffa model construction as well. For this we need enough model theory to get models with automorphisms.)

7.1.3 Mathematics in *NFU*

(NOTE: Counting is so useful that it might show up in the base development.)

We do not start with a clean slate when we consider doing mathematics in *NFU*, because all the mathematics we have done in *TSTU* can be imported. However, the interpretation of *NFU* is different in interesting ways.

The language of *NFU* is larger. Sentences such as $x \in x$ are well-formed as they are not in typed language. Further, a sentence like $V \in V$ which we wrote but construed as a sort of pun in typed language is to be taken seriously in *NFU*: the universal set V has *everything* as an element, including itself. From this it follows that $(\exists x.x \in x)$ is a theorem of *NFU*, since the universal set is a witness.

We have proved Cantor's theorem $|\iota``A| < |\mathcal{P}(A)|$ which tells us that the power set of A is larger than A. But in *NFU* we of course know that $\mathcal{P}(V) \subseteq V$. This does not contradict anything we proved in type theory, because in type theory the referents of the two V's are not supposed to be the same. In *NFU* Cantor's Theorem tells us that $|\iota``V| < |\mathcal{P}(V)| \leq |V|$, so we see that the singleton map (which from an external standpoint we can see is a one-to-one correspondence) cannot be a set in *NFU*.

The unstratified form of Cantor's Theorem which is true in the untyped set theories of chapter 4 cannot hold in general in *NFU*, but it can hold under special circumstances.

Definition: A set A is said to be *cantorian* iff $|A| = |\iota``A|$.

This is precisely what is needed to get the unstratified theorem "if A is a cantorian set, $|A| = |\iota``A| < |\mathcal{P}(A)|$". We see that all cantorian sets are smaller than their power sets. Consideration of how this fact is witnessed suggests a stronger property.

Definition: A set A is said to be *strongly cantorian* iff $(\iota\lceil A) = \{(a, \{a\}) \mid a \in A\}$ is a set.

Obviously a strongly cantorian set is cantorian. The stronger property has considerably stronger consequences.

What all of this already tells us is that a model of *NFU* is not a model of *TSTU* of the natural kind in which every collection of type i objects is a type $i + 1$ object. Every element of the non-function $\iota = \{(x, \{x\}) \mid x \in V\}$ is an object in our model of *NFU*, but the collection of all these pairs cannot be an element of the model on pain of contradiction.

We give a much sharper result of the same kind. We proved above that $T^2(\Omega) < \Omega$ (recall that Ω is the order type of the ordinals). In *TSTU* this assertion was a kind of pun, but here all references to Ω are references to the same object. It is straightforward to prove that $\alpha < \beta \leftrightarrow T(\alpha) < T(\beta)$, from which it follows that $\Omega > T^2(\Omega) > T^4(\Omega) > T^6(\Omega) > \ldots$. This observation has two different rather alarming consequences. One is that a certain *countable* collection of objects of a model of *NFU* cannot be a set: if the smallest collection containing Ω and closed under T^2 were a set, it would be a set of ordinals with no smallest element, which is impossible. The other is that from a certain external standpoint, the ordinals of a model of *NFU* are not well-ordered.

We investigate the mathematics of the properties "cantorian" and "strongly cantorian".

Theorem: Concrete finite sets are cantorian. Power sets of cantorian sets are cantorian. Cartesian products of cantorian sets are cantorian. Function spaces from cantorian sets to cantorian sets are cantorian.

Proof: Sets of concrete finite sizes are obviously the same size as their images under the singleton operation. We will find that asserting this for all finite sets is a stronger assertion than we can prove from our current axioms. The other assertions follow from the existence of bijections between $\mathcal{P}(\iota"A)$ and $\iota"(\mathcal{P}(A))$, between $\iota"A \times \iota"B$ and $\iota"(A \times B)$ and between $\iota"B^{\iota"A}$ and $\iota"(B^A)$: from the ability to define these maps it clearly follows that if A, B are the same size as $\iota"A$, $\iota"B$, respectively, then $\mathcal{P}(A)$, $A \times B$, B^A are the same size as $\iota"\mathcal{P}(A)$, $\iota"(A \times B)$, $\iota"(B^A)$, respectively, which is what is to be shown.

Theorem: Concrete finite sets are strongly cantorian. Power sets of cantorian sets are strongly cantorian. Cartesian products of cantorian sets are strongly cantorian. Function spaces from cantorian sets to cantorian sets are strongly cantorian.

Proof: If A is a concrete finite set, $(\iota\lceil A)$ can be given as a concrete finite set. Again, showing that this is true for all finite sets turns out not to be provable with our current axioms. Construct $(\iota\lceil\mathcal{P}(A))$ as $(B : \mathcal{P}(A) \mapsto (A : \mathcal{P}(\iota\text{“}V) \mapsto \{\bigcup A\})((\iota\lceil A)\text{“}B))$. Construct $(\iota\lceil(A \times B))$ as $((a, b) : A \times B \mapsto ((\{x\}, \{y\}) : (\iota\text{“}V) \times (\iota\text{“}V) \mapsto \{(x, y)\})((\iota\lceil A)(a), (\iota\lceil B)(b)))$. We leave the similar construction of $(\iota\lceil B^A)$ as an exercise.

Theorem: A subset of a strongly cantorian set is strongly cantorian.

Proof: If $B \subseteq A$, $(\iota\lceil B) = (\iota\lceil A)\lceil B$.

The last theorem is one reason why "strongly cantorian" is a much stronger property. Here is a further, more profound reason.

Subversion Theorem: Suppose that for a given formula ϕ there is a function σ from variables appearing in ϕ to integers

7.1.4 There are Urelements

7.2 Extensions of *NFU*

7.2.1 The Axiom of Counting; ω-Models.

But perhaps Counting will be covered in the first part?
unstratified induction? The ω-model construction; α-models; NFU*.

7.2.2 The Axiom of Cantorian Sets; the Axiom of Large Ordinals

this will provide an occasion for T-sequences. Interpretation of *ZFC* in this theory (cute eliminations of T). n-Mahlos, fancy partition relations, model theory.

7.2.3 The Axiom of Small Ordinals; the BEST model

ASO with and without CS and Large Ordinals. weakly compact; nearly measurable. Solovay stuff. The BEST model.

7.3 The Extensional Subsystems

7.3.1 Ambiguity in Theories with Finitely Many Types; NF_3

Our type theory $TSTU$ has natural subtheories defined simply by restricting the number of types. Similar considerations apply to variants of our type theory.

Definition: $TSTU_n$ is defined as the subtheory of $TSTU$ with type indices $\geq n$ excluded from the language. Other type theories will have subscripted variants defined in the same way.

The situation in three types is very special.

Theorem: For any infinite model of $TSTU_3$ with either the same concrete finite number of atoms at each type or infinitely many atoms at each type, there is a model of ${TSTU_3}^\infty$ with exactly the same theory.

Proof: By model theory, there is a countable model of $TSTU_3$ with the same theory. We want a further refinement: we want a countable model with the property that each infinite set can be partitioned into two infinite sets. Suppose our initial countable model lacks this property: there are then infinite sets which can only be partitioned into finite and cofinite pieces. Construct an ultrapower of the model using an ultrafilter on the natural numbers. This will give a model of the theory with the splitting property (but not a countable one). Build a countable model with the same theory as this model, but being sure to include some specific constant (referring to a set of nonstandard finite size) in your theory. The resulting model will be countable, will have the splitting property (because we will have partitions of any infinite set with one partition of the fixed nonstandard size), and will have exactly the same theory as the original model (if we exclude references to the special constant from our language).

Now we show that in any countable model of $TSTU$ there is an isomorphism between types $0-1$ and types $1-2$. First of all, the conditions in the statement of the theorem combined with the countability of the model are enough to ensure that we have a bijection from the type 1 atoms onto the type 2 atoms. Now we handle the sets. We fix an order on the type 1 sets and an order on the type 2 sets, each of type ω. When

we have mapped the first n sets of type 1 to sets of type 2, and also the first n sets of type 2 have been assigned inverse images in type 1, we assume that we have matched them in such a way that the sizes of the corresponding compartments in Venn diagrams determined by the type 1 sets assigned images and the type 2 sets assigned inverse images is correct: for any intersection of the type 1 sets and their complements, if the intersection is of concrete finite size n the corresponding intersection of type 2 sets and their complements will be of the same concrete finite size n, and if the intersection is (countably) infinite the corresponding intersection of the type 2 sets and their complements will be countably infinite. We show how to continue this process (note that the conditions are vacuously satisfied initially). Match the first set of type 1 not yet assigned an image with the first set in the order on type 2 sets which has not yet been matched and has the correct intersection sizes with the correlates of all finite intersections of the previously mapped type 1 sets. The splitting property is needed here to ensure that if the new type 1 set has infinite and co-infinite intersection with one of the compartments of the Venn diagram determined by the previous set that we can choose a type 2 set with appropriate intersection sizes to associate with it. Choose an inverse image for the first type 2 set as yet not assigned an inverse image in exactly the same way. Notice that the map between types 1 and 2 determines a map between types 0 and 1 by considering singletons. Note that the amount of comprehension needed in the type theory considered is very limited: all that is needed is existence of singletons, complements and finite unions.

If f is the isomorphism, we take type 0 as the model and define $x \in_N y$ as $x \in_M f(y)$ (where $\in_M$ is the membership relation of the model. Note that for any $x^{\mathbf{0}} \in_M y^{\mathbf{1}}$ we have $x^{\mathbf{0}} \in_M f^{-1}(y^{\mathbf{1}})$ equivalent and for any $x^{\mathbf{1}} \in_M y^{\mathbf{2}}$ we have $f^{-1}(x^{\mathbf{1}}) \in_M f^{-2}(y^{\mathbf{2}})$ This model N will be a model of $TSTU_3{}^{\infty}$: this should be evident.

It should be evident from these considerations that all models of $TSTU_3$ satisfying the conditions on numbers of atoms (which are describable in terms of sets of sentences satisfied in their theories) also satisfy Amb (noting that the scheme $\phi \leftrightarrow \phi^+$ must be restricted to formulas not mentioning type 2).

Definition: Define NF_3 as the theory whose axioms are Strong Extension-

ality and those instances "$\{x \mid \phi\}$ exists" of Stratified Comprehension which can be stratified using a stratification with range $\{0, 1, 2\}$ (note that the stratification will send x to 0 or 1, since it must assign 1 or 2 to $\{x \mid \phi\}$).

Corollary: NF_3 is consistent.

Proof: In the previous Theorem, fix the number of atoms at 0.

Observation: This is the first consistent fragment of New Foundations which we have identified which has strong extensionality. It is important to notice that, unlike NF, this is *not* a weird theory involving considerations strange to ordinary mathematics. *Every* infinite model of TST_3 has a correlated model of NF_3 which satisfies the same sentences when types are dropped. NF_3, though it may seem unfamiliar, is ubiquitous and should be of considerable interest in foundations of mathematics.

We go on to consider Ambiguity for $TSTU_n$ with $n > 3$.

Theorem: $TSTU_n^\infty$ is consistent iff $TSTU_n + Amb$ is consistent.

Proof: Notice that our proof above depended on being able to iterate the $+$ operation as far as wanted; this is spoiled by the presence of a top type. We will fix this problem using a trick.

We can cleverly delete all reference to the bottom type of our language. We define $[\subseteq]^2$ as the collection of all sets $\{x \mid x \subseteq A\}$ where A is a fixed type 1 set (it is important to recall that an urelement is not a subset of anything). We define $1^{\mathbf{1}}$ as usual as the set of all singletons. We now observe that $x^{\mathbf{0}} \in y^{\mathbf{1}}$ is equivalent to the assertion that $\{x\} \subseteq y$, which is in turn equivalent to "$\{x\}$ belongs to every element of $[\subseteq]^{\mathbf{2}}$ which contains y". We can now replace all references to specific type 0 objects by references to singletons and all quantifiers over type 0 with quantifiers over $1^{\mathbf{1}}$, redefining membership in type 0 objects appropriately.

This doesn't give us anything obvious for free, as we have our special constants 1^1 and $[\subseteq]^2$ to consider. We further observe that it is a theorem that $1^{\mathbf{1}}$ is a subset of the domain of $[\subseteq]^{\mathbf{2}}$ and for every (type 2) subset A of $1^{\mathbf{1}}$ there is a unique type 1 object a in the range of $[\subseteq]^{\mathbf{2}}$

such that "$\{x\} \subseteq a$" (a fact expressible without mentioning type 0) iff $\{x\} \in A$.

Now *Amb* tells us that there are objects $1^{\mathbf{0}}$ and $[\subseteq]^{\mathbf{1}}$ with the type-shifted version of the same property noted above for 1^1 and $[\subseteq]^1$. These can be reinterpreted as the singleton set on a new type -1 and the inclusion relation on type 0 objects construed as "sets" of type -1 objects. This means that $TSTU_n + Amb$ interprets $TSTU_{n+1}$ (we can reindex so that the new type -1 becomes type 0). We can further use ambiguity to ensure that as much as we wish to be true about $1^{\mathbf{0}}$ and $[\subseteq]^{\mathbf{1}}$ is the type-shifted analogue of what is true about $1^{\mathbf{1}}$ and $[\subseteq]^{\mathbf{2}}$ [we cannot show that there are specific relations which have exactly the same properties, merely that there is a relation with any finite selection of type shifted versions of the properties of $1^{\mathbf{1}}$ and $[\subseteq]^{\mathbf{2}}$], and thus show by compactness that the extension of *Amb* can consistently hold as well. So the consistency of $TSTU_n + Amb$ for $n > 3$ implies the consistency of $TSTU_{n+1} + Amb$, whence it implies the consistency of $TSTU + Amb$, whence it implies the consistency of New Foundations.

Corollary: $TST_4 + Amb$ is consistent iff NF is consistent.

Observation: The profound difference between the case $n = 3$ and the case $n = 4$ in the strongly extensional case is of interest here.

Observation: The proofs above will also work in some other type theories.

Make the point that NF style considerations are natural and ubiquitous in 3-typed mathematics.

This should include the proof of consistency of NFU using 3-type machinery and the Pigeonhole Principle instead of Ramsey's theorem.

Mathematics in three types, functions without pairs. FM methods in the first section would avoid an inversion here.

7.3.2 Predicativity; NFP; The Ramified Theory of Types Interpreted in NFP; NFI

7.4 Finite Universes: *NFU* + "the universe is finite".

also NFU and nonstandard analysis?

7.5 New Foundations

7.5.1 History of *NF*; Errors of Quine

Specker trees, all the bad stuff. A section on FM methods in type theory would help here as it would provide an occasion in the first part to carefully discuss choice-free mathematics. Orey's metamathematical results; of course they also work in *NFU*.

7.6 Technical Methods for Consistency and Independence Proofs in *NF(U)*

7.6.1 Forcing in Type Theory and Set Theory

Introduce the method of forcing in *NFU* at least and possibly in type theory and ordinary set theory. Prove the independence of the continuum hypothesis. Forcing in *NF*, of course. But this may continue a section on forcing in the type theory part.

7.6.2 Frankel-Mostowski Permutation Methods

Prove the independence of the Axiom of Choice from type theory (certainly) and possibly from *NFU* and/or ordinary set theory. The initial parts of this may occur in the type theory part.

7.7 Cut Elimination in Type Theory and Set Theory

Prove cut elimination in type theory and *SF*. Maybe other applications of Marcel's weak extensional collapse.

7.8 Stratified Combinatory Logic and λ-Calculus

7.9 Rieger-Bernays Permutation Methods

Explore the consistency and independence proofs obtainable, and the set based notion of "well-foundedness" and related ideas. Unstratified implementations of numerals.

7.10 Limitations of Universal Constructions

The existence of universal objects is not magic. Cartesian closedness failing for the category of all sets and functions is an advantage.

8 Philosophy of Set Theory

General considerations about the relative merits of the various systems considered here and about the sufficiency of each as a foundational system. Comments on the general weirdness of *NF* and the real nature of the *NF* consistency problem belong here.